THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XX

1949

CONTENTS OF VOLUME XX

Articles

Anderson, T. W. and Herman Rubin. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations . 46

Andrews, F. C. and Z. W. Birnbaum. On Sums of Symmetrically Truncated Normal Random Variables. 458

Baker, G. A. The Variance of the Proportions of Samples Falling Within a Fixed Interval for a Normal Population. 123

Banerjee, K. S. A Note on Weighing Design. 300

Bancroft, T. A. Some Recurrence Formulae in the Incomplete Beta Function Ratio. 451

Barankin, E. W. Locally Best Unbiased Estimates. 477

Berger, Agnes and Abraham Wald. On Distinct Hypotheses. 104

Birnbaum, Z. W. and F. C. Andrews. On Sums of Symmetrically Truncated Normal Random Variables. 458

Birnbaum, Z. W. and H. S. Zuckerman. A Graphical Determination of Sample Size for Wilks' Tolerance Limits 313

Blom, Gunnar. A Generalization of Wald's Fundamental Identity. 439

Boas, R. P., Jr. Representation of Probability Distributions by Charlier Series 376

Bose, R. C. A Note on Fisher's Inequality for Balanced Incomplete Block Designs 619

Chernoff, Herman. Asymptotic Studentization in Testing of Hypotheses. 268

David, Herbert T. A Note on Random Walk. 603

Doob, J. L. Heuristic Approach to the Kolmogorov-Smirnov Theorems. . 393

Dvoretzky, Aryeh. On the Strong Stability of a Sequence of Events. . . 296

Dwyer, Paul S. Pearsonian Correlation Coefficients Associated with Least Squares Theory. 404

Epstein, Benjamin. A Modified Extreme Value Problem 99

Epstein, Benjamin. The Distribution of Extreme Values in Samples Whose Members are Subject to a Markoff Chain Condition 590

Erdös, P. On a Theorem of Hsu and Robbins. 286

Godwin, H. J. A Note on Kac's Derivation of the Distribution of the Mean Deviation. 127

Godwin, H. J. Some Low Moments of Order Statistics. 279

Goodman, Leo A. On the Estimation of the Number of Classes in a Population. 572

Greenwood, Robert E. Numerical Integration for Linear Sums of Exponential Functions. 608

Grubbs, Frank E. On Designing Single Sampling Inspection Plans 242

Halmos, Paul R. and L. J. Savage. Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics 225

HANSEN, MORRIS H. AND WILLIAM N. HURWITZ. On the Determination of Optimum Probabilities in Sampling 426
HATKE, SISTER MARY AGNES. A Certain Cumulative Probability Function 461
HOEL, P. G. AND R. P. PETERSON. A Solution to the Problem of Optimum Classification 433
HORTON, H. BURKE AND R. TYNES SMITH, III. A Direct Method for Producing Random Digits in Any Number System 82
HOWELL, JOHN M. Control Chart for Largest and Smallest Values 305
HURWITZ, WILLIAM N. AND MORRIS H. HANSEN. On the Determination of Optimum Probabilities in Sampling 426
KIMBALL, BRADFORD F. An Approximation to the Sampling Variance of an Estimated Maximum Value of Given Frequency Based on Fit of Doubly Exponential Distribution of Maximum Values 110
LEHMANN, E. L. AND C. STEIN. On the Theory of Some Non-Parametric Hypotheses 28
LEV, JOSEPH. The Point Biserial Coefficient of Correlation 125
LEVENE, HOWARD. On a Matching Problem Arising in Genetics 91
MATERN, BERTIL. Independence of Non-Negative Quadratic Forms in Normally Correlated Variables 119
MADOW, WILLIAM G. On the Theory of Systematic Sampling, II 333
MCMILLAN, BROCKWAY. Spread of Minima of Large Samples 444
MOOD, A. M. Tests of Independence in Contingency Tables as Unconditional Tests 114
NOETHER, GOTTFRIED E. On a Theorem by Wald and Wolfowitz 455
OLDS, EDWIN G. The 5% Significance Levels for Sums of Squares of Rank Differences and a Correction 117
OTTER, RICHARD. The Multiplicative Process 206
PAULSON, EDWARD. A Multiple Decision Procedure for Certain Problems in the Analysis of Variance 95
PEISER, A. M. Correction to "Asymptotic Formulas for Significance Levels of Certain Distributions" 128
PETERSON, R. P. AND P. G. HOEL. A Solution to the Problem of Optimum Classification 433
PITMAN, E. J. G. AND HERBERT ROBBINS. Application of the Method of Mixtures to Quadratic Forms in Normal Variates 552
QUENOUILLE, M. H. Problems in Plane Sampling 355
QUENOUILLE, M. H. The Joint Distribution of Serial Correlation Coefficients 561
REICH, EDGAR. On the Convergence of the Classical Iterative Method of Solving Linear Simultaneous Equations 448
RIORDAN, JOHN. Inversion Formulas in Normal Variable Mapping 417
ROBBINS, HERBERT AND E. J. G. PITMAN. Application of the Method of Mixtures to Quadratic Forms in Normal Variates 552

RUBIN, HERMAN AND T. W. ANDERSON. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations 46
SAVAGE, L. J. AND PAUL R. HALMOS. Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics 225
SARD, ARTHUR. Smoothest Approximation Formulas 612
SETH, G. R. On the Variance of Estimates 1
SOBEL, MILTON AND ABRAHAM WALD. A Sequential Decision Procedure for Choosing one of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution 502
SMITH, R. TYNES, III AND H. BURKE HORTON. A Direct Method for Producing Random Digits in Any Number System 82
STEIN, C. AND E. L. LEHMANN. On the Theory of Some Non-Parametric Hypotheses 28
TUKEY, JOHN W. Sufficiency, Truncation and Selection 309
TUKEY, JOHN W. Moments of Random Group Size Distributions 523
VON SCHELLING, HERMANN. A Formula for the Partial Sums of Some Hypergeometric Series. 120
WALD, ABRAHAM AND AGNES BERGER. On Distinct Hypotheses 104
WALD, ABRAHAM. Statistical Decision Functions 165
WALD, ABRAHAM. Note on the Consistency of the Maximum Likelihood Estimate 595
WALD, ABRAHAM AND MILTON SOBEL. A Sequential Decision Procedure for Choosing one of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution 502
WALSH, JOHN E. Some Significant Tests for the Median which are Valid Under Very General Conditions 64
WALSH, JOHN E. On the Range-Midrange Test and Some Tests with Bounded Significance Levels 257
WALSH, JOHN E. On the Power Function of the "Best" t-test Solution of the Behrens-Fisher Problem 616
WALSH, JOHN E. Concerning Compound Randomization in a Binary System 580
WOLFOWITZ, J. On Wald's Proof of the Consistency of the Maximum Likelihood Estimate 601
WOLFOWITZ, J. The Power of the Classical Tests Associated with the Normal Distribution 540
WOODBURY, MAX A. On a Probability Distribution 311
YOSIDA, KÔSAKU. Brownian Motion on the Surface of the 3-Sphere 292
ZUCKERMAN, H. S. AND Z. W. BIRNBAUM. A Graphical Determination of Sample Size for Wilks' Tolerance Limits 313

MISCELLANEOUS

Abstracts of Papers 130, 317, 464, 620
Constitution and By-Laws of the Institute 327

Election of Officers and Council and Revision of By-Laws 150
News and Notices 141, 321, 470, 624
Report on the Berkeley Meeting of the Institute 475
Report on the Boulder Meeting of the Institute 626
Report on the Cleveland Meeting of the Institute 152
Report on the New York Meeting of the Institute 325
Report on the Seattle Meeting of the Institute 151
Report of the President of the Institute 156
Report of the Secretary-Treasurer of the Institute 160
Report of the Editor 163

ON THE VARIANCE OF ESTIMATES

By G. R. Seth

Columbia University

Summary. In this paper recent results on the lower bound to the variance of unbiased estimates have been brought together. Some of them have been extended to sequential estimates and the others have been improved to some extent. In the last section a general method for generating a system of orthogonal polynomials with respect to a certain class of weight functions is obtained together with a result on the conditions under which the class of unbiased estimates formed by all functions of an unbiased estimate consists of just one element.

1. Introduction.

§1.1. Let $X_1, X_2 \cdots$ be a sequence of chance variables whose distribution depends upon an unknown parameter θ and possibly also a finite number of other parameters. It is assumed that either all the X's are absolutely continuous or that they are all discrete. Let $p_M(x_1, x_2, \cdots, x_M; \theta)$ denote the joint probability density function or the probability of $(X_1, \cdots, X_M)$ according as the X's are continuous or discrete. Let $\theta^*(x_1, x_2, \cdots, x_n)$ be an unbiased estimate of θ, where $x_1, x_2, \cdots, x_n$ is a sequence of observations on $X_1, X_2, \cdots, X_n$.

In this paper, we shall make use of the following short forms and abbreviations:

$E(X)$ will represent the expectation of X.

$\sigma^2(X)$ will represent the variance of X.

$E(y \mid x)$ will represent the conditional expectation of y, given x.

θ^* will represent an abbreviation of $\theta^*(x_1, x_2, \cdots, x_n)$.

f will represent an abbreviation of $f(x; \theta)$ or $f(x; \theta_1, \theta_2, \cdots, \theta_T)$.

p_n will represent an abbreviation of $p_n(x_1, x_2, \cdots, x_n; \theta)$ or $p_n(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_T)$.

p_N will represent p_n for a fixed size sample, i.e., $n = N$.

g will represent an abbreviation of $g(\theta^*; \theta)$ or $g(\theta_1^*, \theta_2^*, \cdots, \theta_T^*; \theta_1, \theta_2, \cdots, \theta_T)$.

h will represent $h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta)$ or $h(\xi_1, \xi_2, \cdots, \xi_{N-T} \mid \theta_1^*, \theta_2^*, \cdots, \theta_T^*; \theta_1, \theta_2, \cdots \theta_T)$.

$\phi_{i_1, i_2, \cdots, i_T}(n)$ will represent $\dfrac{1}{p_n} \cdot \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} p_n$.

$g_{i_1, i_2, \cdots, i_T}$ will represent $\dfrac{1}{g} \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} g$.

$h_{i_1, i_2, \cdots, i_T}$ will represent $\dfrac{1}{h} \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} h$.

In case differentiations with respect to one parameter are involved, the last three abbreviations will be shortened to $\phi_i(n)$, g_i and h_i respectively.

In §1.1, n is assumed to be a constant equal to N, that is, the sequence of chance variables is finite and fixed, consisting of $X_1, X_2, X_3, \cdots, X_N$.

Cramer [1] and Rao [2] have shown that under certain conditions of regularity, the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ satisfies the inequality:

$$\sigma^2\theta^*(x_1, x_2, \cdots, x_N) \geq \frac{1}{E\left(\frac{1}{p_N}\frac{\partial p_N}{\partial \theta}\right)^2}. \tag{1.1 1}$$

Cramér [1] has shown that the lower bound for the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ given by (1.1.1) is achieved if and only if:

(1.1.2). There exists a sufficient statistic for estimating θ.

(1.1.3). The probability distribution $g(\theta^*; \theta)$ of the sufficient statistic $\theta^*(x_1, x_2, \cdots, x_N)$ is of the form

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = \frac{K}{g(\theta^*; \theta)}\frac{\partial}{\partial\theta}g(\theta^*; \theta), \quad \text{whenever} \quad g(\theta^*; \theta) > 0,$$

where K depends only upon N and the parameters in the distribution.

Cramer calls the statistic $\theta^*(x_1, x_2, \cdots, x_N)$ satisfying (1.1.2) and (1.1.3) an "efficient" statistic estimating θ and we will use the word "efficient" in this sense alone Bhattacharyya [3] has shown that there exists a lower bound to the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ which is higher than or equal to the one given in (1.1.1). This lower bound is ${}_{(m)}\lambda^{11}$, that is,

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_N)) \geq {}_{(m)}\lambda^{11} \tag{1.1.4}$$

where

$$\| {}_{(m)}\lambda^{ij} \| = \| \lambda_{ij} \|^{-1},$$

and

$$\lambda_{ij} = E\left(\frac{1}{p_N^2}\frac{\partial^i p_N}{\partial\theta^i}\frac{\partial^j p_N}{\partial\theta^j}\right), \qquad i, j = 1, 2, \cdots, m, \tag{1.1.5}$$

where m is any positive integer.

Let θ consist of T components $\theta_1, \theta_2, \cdots, \theta_T$, and $p_N(x_1, x_2, \cdots, x_N; \theta_T)$ be the same as $\prod_{i=1}^{N} f(x_i, \theta_1, \theta_2, \cdots, \theta_T)$. Further let $\theta_1^*(x_1, x_2, \cdots, x_N)$, $\theta_2^*(x_1, x_2, \cdots, x_N), \cdots, \theta_T^*(x_1, x_2, \cdots, x_N)$ be unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ respectively, with the non-singular covariance matrix $\| V_{ij} \|$ $(i, j = 1, 2, \cdots, T)$. Cramér [4] has proved that under certain regularity conditions, the ellipsoid

$$\sum_{i,j=1}^{T} V^{ij}t_i t_j = T + 2 \tag{1 1.6}$$

contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} I_{ij}t_i t_j = T + 2, \tag{1.1.7}$$

where

$$\| V^{ij} \|^{-1} = \| V_{ij} \|, \tag{1.1.8}$$

and

$$I_{ij} = E\left(\frac{N}{f^2}\frac{\partial f}{\partial \theta_i}\cdot\frac{\partial f}{\partial \theta_j}\right). \tag{1.1.9}$$

This result is also implicitly contained in Rao [2].

§1.2. Let us now take n as a chance variable determined by a sequential procedure. $X_1, X_2, X_3, \cdots$ is a sequence of chance variables having the same probability density or probability $f(x;\theta)$, according as X is absolutely continuous or discrete. The sequential process tells us, after each successive observation has been drawn, whether the next observation is to be taken or not. Thus n will denote the total number of observations taken by the time the sequential process has been completed. Under certain regularity conditions, Wolfowitz [5] has shown that if $\theta^*(x_1, x_2, \cdots, x_n)$ is an unbiased estimate of θ, then

$$\sigma^2\theta^*(x_1, x_2, \cdots, x_n) \geq \frac{1}{En\cdot E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2}. \tag{1.2.1}$$

Furthermore, if θ consists of T components, $\theta_1, \theta_2, \cdots, \theta_T$, and $\theta_1^*(x_1, x_2, \cdots, x_n)$, $\theta_2^*(x_1, x_2, \cdots, x_n), \cdots, \theta_T^*(x_1, x_2, \cdots, x_n)$ are unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ respectively, Wolfowitz [5] has proved that

$$\sum_{i,j=1}^{T} I_{ij}t_it_j = T + 2 \tag{1.2.2}$$

is contained within the ellipsoid

$$\sum_{i,j=1}^{T} V^{ij}t_it_j = T + 2, \tag{1.2.3}$$

where

$$I_{ij} = En\cdot E\left(\frac{\partial \log f}{\partial\theta_i}\frac{\partial \log f}{\partial\theta_j}\right), \qquad i,j = 1, \cdots, T.$$

Blackwell and Girshick [6] have shown that the lower bound given by (1.2.1) for the variance of an unbiased estimate of θ is attained only for the sequential process for which $\Pr(n = N) = 1$, if the probability density function $f(x;\theta)$ of X is such that $E(X) = \theta$ and $x_1 + x_2 + x_3, \cdots + x_M$ is a sufficient statistic for all integral values of M, for estimating θ; $x_1, x_2, \cdots, x_M$ being M independent observations on the chance variable X.

In this paper the following results have been obtained. The specific conditions under which the results hold are stated at their proper places along with the results:

(1.3.1) The lower bound in (1.1.4) is valid when n is considered a

chance variable determined by a sequential procedure instead of being a fixed number N.

(1.3.2) The concentration ellipsoid defined in (1.2.3) contains within itself another ellipsoid

$$\sum_{i,j=1}^{r} \mu_{ij} t_i t_j = T + 2$$

where μ_{ij} is given by (3.1.18), which in turn contains the ellipsoid given by (1.2.2).

(1.3.3). The Blackwell and Girshick result [6] for the achievement of the lower bound for the variance of unbiased estimates given by (1.2.1) has been extended to the case where the probability density (or probability) $\prod_{i=1}^{M} f(x_i ; \theta)$, for all fixed $M \geq N$, where N is the least value for which $\Pr(n = N) \neq 0$, has an unbiased "efficient" estimate for θ in the sense defined by Cramer. This is illustrated by two examples of Wald sequential procedures.

(1.3.4). Let N be fixed and $p_N(x_1, x_2, \cdots, x_N ; \theta) \cdot |J| = g(\theta^*; \theta) h(\xi_1, \xi_2, \cdots, \xi_{N-1} | \theta^*, \theta)$, where J denotes the Jacobian of the transformation from $x_1, x_2, \cdots, x_N$ to $\theta^*, \xi_1, \xi_2, \cdots, \xi_{N-1}$. Here $g(\theta^*; \theta)$, and $h(\xi_1, \xi_2, \cdots, \xi_{N-1} | \theta^*; \theta)$ are respectively the probability density function (or probability) of θ^* and the conditional probability density function (or probability) of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for a given value of θ^*.

The necessary and sufficient conditions under which the lower bound for the variance of unbiased estimates given by Bhattacharyya [3] may be achieved are that there should exist a statistic $\theta^*(x_1, x_2, \cdots, x_N)$ such that:

(a) $h_1, h_2, \cdots, h_m$ are linearly dependent considered as functions of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for given values of θ and $\theta^*(x_1, x_2, \cdots, x_N)$ and

(b) the probability density $g(\theta^*; \theta)$ of $\theta^*(x_1, x_2, \cdots, x_N)$ satisfies the following equation:

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = \sum_{i=1}^{n} \frac{K_i}{g(\theta^*; \theta)} \frac{\partial^i}{\partial \theta^i} g(\theta^*; \theta),$$

where K_i are independent of the $x_1, x_2, x_3, \cdots, x_N$.

Equivalent conditions for the multiparameter case have also been given.

(1.3.5). The following properties of $\phi_1(n), \phi_2(n), \cdots$ are derived:

(a) Under certain conditions $\phi_1(N), \phi_2(N) \cdots$ form a system of orthogonal polynomials in $\phi_1(N)$, the weight function being $p_N(x_1, x_2, \cdots, x_N ; \theta)$.

(b) $\sum_{i=1}^{m} K_i \phi_i(n)$ cannot be a function of $x_1, x_2, \cdots, x_n$, independent of θ except for the constant zero.

(c) If $\theta^*(x_1, x_2, \cdots, x_n)$ is linearly dependent upon $\phi_1(n)$, then no other statistic except of the form $a\theta^*(x_1, x_2, \cdots, x_n) + b$ where a and b are constant independent of θ, can be linearly related with $\phi_1(n)$.

(1.3.6). If a) $\theta^*(x_1, x_2, \cdots, x_N)$ is an unbiased estimate of θ and b) if among

all functions of $\theta^*(x_1, x_2, \cdots, x_N)$ which are unbiased estimates of θ with finite variance, θ^* is the one with the least variance and such that the set of polynomials with respect to the distribution function of θ^* is complete, then there is no function of θ^* having a finite variance which is an unbiased estimate of θ.

2. Estimation of a single parameter.

§2.1. Let $X_1, X_2, \cdots$ and $p_M(x_1, x_2, \cdots x_M; \theta)$ be as given in the first paragraph of (1.1). Let Ω be the space of all possible infinite sequences (ω) of observations $x_1, x_2, \cdots$. Let there be given an infinite sequence of Borel measurable functions $\Phi_1(x_1), \Phi_2(x_1, x_2), \cdots, \Phi_j(x_1, x_2, x_3, \cdots, x_j) \cdots$, defined for all observable sequences in Ω such that each takes only the values zero and one. We further assume that everywhere in Ω, except possibly on a set whose probability is zero for all θ under consideration at least one of the functions $\Phi_1(x_1)$, $\Phi_2(x_1, x_2), \cdots$ takes the value of one. Let n be the smallest integer for which this occurs. Thus $n(\omega)$ is a chance variable. The sequential process is then defined as follows:

Take an observation and find $\Phi_1(x_1)$. If it is unity, the sampling process stops; otherwise continue sampling If a second observation is taken and the value of $\Phi_2(x_1, x_2)$ is unity, the process stops; otherwise continue sampling, and so on. In general, if after taking j observations

$$\Phi_i(x_1, x_2, \cdots, x_i) = 0 \text{ for } i = 1, 2, \cdots j - 1,$$

and $\Phi_j(x_1, x_2, \cdots, x_j) = 1$, sampling stops; otherwise it is continued. We will denote by R_j the set of all points $(x_1, x_2, \cdots, x_j)$ for which the process stops with the jth observation

Let $\theta^*(x_1, x_2, \cdots, x_n)$ be a statistic whose expectation is a real valued function $\gamma(\theta)$ of θ. The development proceeds on the assumption that $p_M(x_1, x_2, \cdots, x_M; \theta)$ is a probability density function. The result is equally valid if $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability of discrete variables X_1, $X_2, \cdots, X_M$ provided that integration is replaced by summation whenever this is required. Further the phrase "almost all points" in a Euclidean space of any finite dimensionality is understood to mean all points in the space with the following possible exceptions:

(a). A set of Lebesgue measure zero where $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability density function;

(b). The points which belong to the set Z, where $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability function of the discrete chance variables $X_1, X_2, \cdots, X_M$. The set Z consists of all points $(x_1, x_2, \cdots, x_M)$ such that $p_M(x_1, x_2, \cdots, x_M; \theta) = 0$ identically for all θ under consideration.

§2.2. *Conditions of regularity.* We will postulate the following conditions to be satisfied by $p_M(x_1, x_2, \cdots, x_M; \theta)$ and $\theta^*(x_1, x_2, \cdots, x_n)$.

(2.2.1) $\theta^*(x_1, x_2, \cdots x_n)$ has an expectation $\gamma(\theta)$ and a finite variance. All the derivations of $\gamma(\theta)$ are assumed to be finite. The parameter θ lies in an open interval D of the real line D may consist of the entire line or an entire half line.

(2.2.2). The derivatives

$$\frac{\partial^i p_M}{\partial \theta^i}, \qquad (i = 1, 2, \cdots, m),$$

exist for all θ in D and almost all $x_1, x_2, \cdots, x_M$ in R_M and for all M. We define

$$\frac{1}{p_M}\frac{\partial^i p_M}{\partial \theta^i} = 0,$$

whenever $p_M(x_1, x_2, \cdots, x_M; \theta) = 0$; thus,

$$\frac{1}{p_M}\frac{\partial^i p_M}{\partial \theta^i} = \phi_i(M)$$

is defined for all θ in D and almost all $(x_1, x_2, \cdots, x_M)$ in R_M.

(2.2.3). For any integral j there exists non-negative L-measurable functions $T_i(x_1, x_2, \cdots, x_j)$, $(i = 1, 2, \cdots, m)$, such that

(a) $$| \theta^*(x_1, x_2, \cdots, x_j)\frac{\partial^i}{\partial \theta^i} p_j(x_1, x_2, \cdots, x_j; \theta) | < T_i(x_1, x_2, \cdots, x_j),$$

for all θ in D and almost all $(x_1, x_2, \cdots, x_j)$ in R_j.

(b) $$\int_{R_j} T_i(x_1, x_2, \cdots, x_j) \sum_{u=1}^{j} dx_u, \quad (i = 1, 2, \cdots, m),$$

are finite.

(2.2.4). Let $t_j(\theta) = \int_{R_j} \theta^*(x_1, x_2, \cdots, x_j) p_j(x_1, x_2, \cdots, x_j; \theta) \prod_{u=1}^{j} dx_u$.

We postulate the uniform convergence of

$$\sum_{j=1}^{\infty} \frac{d^i}{d\theta^i} t_j(\theta), \qquad (i = 1, 2, \cdots, m)$$

(the existence of $\frac{d^i}{d\theta^i}(t_j(\theta))$ is assured by the assumption (2.2.3).)

(2.2.5). There exist functions $S_i(x_1, x_2, \cdots, x_j)$ for every j, $(i = 1, 2, \cdots, m)$, such that when $\theta^*(x_1, x_2, \cdots, x_j)$ and $T_i(x_1, x_2, \cdots, x_j)$ are replaced by unity and $S_i(x_1, x_2, \cdots, x_j)$ respectively, conditions (2.2.3) and (2.2.4) still hold good.

(2.2.6). The covariance matrix of $\phi_i(n)$ $(i = 1, \cdots, m)$ exists and is non-singular for almost all θ in D and almost all $(x_1, x_2, \cdots, x_n)$.

§2.3. Let us consider the sequential process mentioned in §2.1 and the functions $\theta^*(x_1, x_2, \cdots, x_n)$ and $p_M(x_1, x_2, \cdots, x_M; \theta)$ which satisfy the regularity conditions in §2.2. We will now find a lower bound for the variance of such estimates.

Let us examine

$$F = E\left(\theta^*(x_1, x_2, \cdots x_n) - \gamma(\theta) - \sum_{i=1}^{m} K_i \phi_i(n)\right)^2, \tag{2.3.1}$$

where K_i $(i = 1, 2, \cdots, m)$ are independent of $(x_1, x_2, \cdots x_n)$. Now (2.3.1) can be written as

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - 2\sum_{i=1}^{m} K_i E\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n) + 2\gamma(\theta)\sum_{i=1}^{m} K_i E\phi_i(n) + \sum_{i,j=1}^{m} K_i K_j \lambda_{ij}, \tag{2.3.2}$$

where

$$\lambda_{ij} = E(\phi_i(n)\phi_j(n)) \qquad (i, j = 1, 2, \cdots, m).$$

Now

$$E(\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n)) = \sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, x_2, \cdots, x_j) \frac{\partial^i p_j}{\partial\theta^i} \prod_{u=1}^{j} dx_u. \tag{2.3.4}$$

We also know that

$$\sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, x_2, \cdots, x_j) p_j \prod_{u=1}^{j} dx_u = \gamma(\theta). \tag{2.3.5}$$

Differentiating both sides of (2.3.5) i times $(i = 1, 2, \cdots, m)$ we have, because of conditions (2.2.3) and (2.2.4):

$$\sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, x_2, \cdots, x_j) \frac{\partial^i p_j}{\partial\theta^i} \prod_{u=1}^{j} dx_u = \frac{d^i\gamma(\theta)}{d\theta^i}, \qquad (i = 1, 2, \cdots, m). \tag{2.3.6}$$

From (2.3.4) and (2.3.6), we obtain

$$E(\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n)) = \frac{d^i}{d\theta^i}\gamma(\theta). \tag{2.3.7}$$

Differentiating

$$1 = \sum_{j=1}^{\infty} \int_{R_j} p_j \prod_{u=1}^{j} dx_u \tag{2.3.8}$$

i times $(i = 1, 2, \cdots, m)$ with respect to θ, we obtain because of conditions (2.2.5)

$$0 = \sum_{j=1}^{\infty} \int_{R_j} \frac{\partial^i p_j}{\partial\theta^i} \prod_{u=1}^{j} dx_u, \qquad (i = 1, \cdots, m). \tag{2.3.9}$$

(2.3.8) is valid on account of the type of sequential process (2.1). Now

$$E(\phi_i(n)) = \sum_{j=1}^{\infty} \int_{R_j} \frac{\partial^i p_j}{\partial\theta^i} \prod_{u=1}^{j} dx_u, \qquad (i = 1, \cdots, m). \tag{2.3.10}$$

By (2.3.7) and (2.3.10), (2.3.2) reduces to

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - 2\sum_{i=1}^{m} K_i \frac{d^i\gamma(\theta)}{d\theta^i} + \sum_{i,j=1}^{m} K_i K_j \lambda_{ij}. \tag{2.3.11}$$

Now $||\lambda_{ij}||$ being non-singular on account of condition (2.2.6), we get just one set of values of K's which minimize F. These values are given by

$$K_j = \sum_{i=1}^{m} {}_{(m)}\lambda^{ij} \frac{d^i\gamma(\theta)}{d\theta^i}, \tag{2.3.12}$$

where

$$\|{}_{(m)}\lambda^{ij}\|^{-1} = \|\lambda_{ij}\|, \qquad (i, j = 1, 2, \cdots, m) \tag{2.3.13}$$

Putting the above values of $K_j (j = 1, 2, \cdots, m)$ in (2.3.11), we obtain

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - \sum_{i,j=1}^{m} {}_{(m)}\lambda^{ij} \cdot \frac{d^i\gamma(\theta)}{d\theta^i} \cdot \frac{d^j\gamma(\theta)}{d\theta^j}. \tag{2.3.14}$$

Hence, F being non-negative by (2.3.1), we have

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq \sum_{i,j=1}^{m} {}_{(m)}\lambda^{ij} \cdot \frac{d^i\gamma(\theta)}{d\theta^i} \cdot \frac{d^j\gamma(\theta)}{d\theta^j}. \tag{2.3.15}$$

Thus R.H.S. of the above inequality gives the lower bound to the variance of unbiased estimates of $\gamma(\theta)$ [1] When $\gamma(\theta) = \theta$, the above reduces to

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq {}_{(m)}\lambda^{11}. \tag{2.3.16}$$

When $m = 1$ and $p_n(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$, (2.3.16) reduces to

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq \frac{1}{En \cdot E\left(\left(\frac{\partial}{\partial\theta} \log f(x; \theta)\right)^2\right)}, \tag{2.3.17}$$

which is the result given by Wolfowitz [5].

When n, the chance variable, is constant and equal to N, then (2.3.15) and (2.3.16) correspond to those given by Bhattacharyya [3]. Although the conditions of regularity under which Bhattacharyya proves his results are not clear from his paper, they are likely to be slightly different from those in §2.3, as the results in [3] are obtained only for a fixed size sample.

§2.4. We will now investigate the necessary and sufficient conditions under which the lower bound given in (2.3.16) is actually higher than that given in (2.3.17).

We can easily see that

$${}_{(m)}\lambda^{11} = \frac{1}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots m})}, \tag{2.4.1}$$

where $R_{1\cdot 23\cdots m}$ is the multiple correlation coefficient between $\phi_1(n)$ and $\phi_2(n)$, $\phi_3(n)$, $\cdots$, $\phi_m(n)$.

The excess of the lower bound given by (2.3.16) over that when we use $m = 1$ is given by

$$\frac{1}{\lambda_{11}(1 - R^2_{1\,23\cdots m})} - \frac{1}{\lambda_{11}} \tag{2.4.2}$$

[1] Under certain weak restrictions, an optimum lower bound to the variance of unbiased estimates has been obtained by me along the lines of a similar result for fixed size samples in an unpublished paper by A. Wald. Independently C Stein has obtained the same result in a paper not yet published

which is further equal to

$$\frac{R^2_{1\,23\cdots m}}{\lambda_{11}(1 - R^2_{1\,23\cdots m})}. \tag{2.4 3}$$

Thus the lower bound for the variance of unbiased estimates of θ is obtained by using $m > 1$ is higher than that obtained by employing $m = 1$ if and only if $R_{1\,23\cdots m}$ is not zero for some $m \geq 2$. This is equivalent to the condition that for at least one $i \geq 2$, λ_{1i}, the correlation coefficient between $\phi_1(n)$ and $\phi_i(n)$ $(i > 1)$, is different from zero. Suppose further that we have used $m = \alpha$ and that we wish to find the increase in the lower bound if α were replaced by $\alpha + 1$. The increase in this case is given by

$$\frac{\rho^2_{1(\alpha+1)\cdot 23\cdots\alpha}}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots(\alpha+1)})} \tag{2.4.4}$$

where $\rho_{1(\alpha+1)\cdot 23\cdots\alpha}$ is the partial correlation coefficient between $\phi_1(n)$ and $\phi_{\alpha+1}(n)$ keeping $\phi_2(n), \cdots, \phi_\alpha(n)$ fixed. It is greater than zero if and only if $\rho_{1(\alpha+1)\cdot 23\cdots\alpha}$ is not equal to zero.

§2 5. If $p_n(x_1, x_2, \cdots, x_n, \theta_1)$ also depends upon a finite number of other parameters $\theta_2, \theta_3, \cdots, \theta_r$, then a lower bound higher than or equal to that given in (2 3.16) can be obtained by using

$$\sum_{i_1+i_2+\cdots+i_r \leq m} K_{i_1,i_2\cdots i_r}\,\phi_{i_1,i_2\cdots i_r}(n) \quad \text{instead of} \quad \sum_{i_1=1}^{m} K_{i_1}\cdot\phi_{i_1}(n) \quad \text{in (2·3·1).}$$

The lower bound in this case is given by (3.1.14) (see section 3) by taking $s = 1$, that is,

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq C(1, 1) \tag{2.5 1}$$

where $C(1, 1)$ is the element in the first row and first column of the inverse of W defined in (3.1.9).

The result for $n = N$, N fixed, is obtained by Bhattacharyya [3, 1947]. Let us illustrate it by an example Take samples of fixed size N. Suppose we are required to find the lower bound to the variance of unbiased estimates of θ_1 in the normal population

$$f(x, \theta_1\,\theta_2) = \frac{1}{\sqrt{2\pi\theta_1}}\; e^{-(x-\theta_2)^2/2\theta_1} \tag{2 5.2}$$

on the basis of N independent observations $x_1, x_2, \cdots, x_N$. The lower bound for the variance of the unbiased estimates of θ_1, when we use

$$\sum_{i_1=1}^{m} K_{i_1}\cdot\phi_{i_1}(N) \text{ in (2 3.1)} \qquad \text{is given by } \frac{2\theta_1^2}{N}.$$

However, if $\sum_{i_1+i_2\leq 2} K_{i_1,i_1}\cdot\phi_{i_1,i_2}(N)$ is used, the lower bound, by the help of (2 5.1), is found to be equal to $2\theta_1^2/(N-1)$ In fact there exists the statistic

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

whose variance is equal to $2\theta_1^2/(N-1)$ where $\bar{x} = \sum_{i=1}^{N} x_i/N$. Thus the use of $\sum_{i_1+i_2\leq 2} K_{i_1,i_2} \cdot \phi_{i_1,i_2}(N)$ brings into relief the unbiased estimate with the least variance.

3. Multi-parameter case. In this section we will prove the result mentioned in (1.3.2) of §1.3.

§3.1. *Let θ consist of T components $(\theta_1, \theta_2, \cdots, \theta_T)$ and $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ be unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ respectively. Also, let a sequential process of the type described in §2.1 be given. We postulate the following regularity conditions:*

(3.1.1). *The covariance matrix $\| V_{ij} \|$ of the estimates $\theta_i^* (i = 1, 2, \cdots, T)$ is non-singular in D, where D is an open interval of the T-dimensional parameter space.*

(3.1.2). *The conditions of section (2.2) are satisfied for each one of $\theta_i^* (i = 1, 2, \cdots, T)$ and $\phi_{i_1,i_2,\ldots,i_T}(n)$, $(i_1 + i_2 + \cdots + i_T \leq m)$.*

(3.1.3). *The covariance matrix of $\phi_{i_1,i_2,\ldots,i_T}(n)$, $i_1 + i_2 + \cdots + i_T \leq m$ exists and is non-singular. Under the assumptions (3.1.1)–(3.1.3), we prove the result (1.3.2) in section 1.3.*

PROOF: Using the same arguments of §2.3, we obtain

$$E(\theta_j^*(x_1, x_2, \cdots, x_n) \cdot \phi_{i_1,i_2,\ldots,i_T}(n)) = 1, \quad \begin{cases} i_\beta = \delta_{\beta j} (\beta = 1, 2, \cdots, T), \\ j = 1, 2, \cdots, T \end{cases} \tag{3.1.4}$$

$$= 0 \text{ otherwise.} \tag{3.1.5}$$

Let the covariance matrix of $\theta_j^* (j = 1, 2, \cdots, s; s \leq T)$ and $\phi_{i_1,i_2,\ldots,i_T}(n)$, $(i_1 + i_2 + \cdots + i_T \leq m)$ be given by

$$U = \begin{Vmatrix} A & B \\ B' & W \end{Vmatrix} \tag{3.1.6}$$

where

$$A = \| V_{ij} \|, \; i, j = 1, 2, \cdots, s; \qquad s \leq T; \tag{3.1.7}$$

$$B = \| I, 0 \|; \tag{3.1.8}$$

(3.1.9) and W = covariance matrix of the set

$$[\phi_{i_1,i_2,\ldots,i_T}(n); i_1 + i_2 + \cdots + i_T \leq m],$$

arranged such that the jth term in the leading diagonal is given by

$$E(\phi_{i_1,i_2,\ldots,i_T}^2(n)), \qquad \text{where } i_j = 1, i_\beta = 0, \beta \neq j, \quad (j = 1, 2, \cdots, T), \tag{3.1.10}$$

and B' is the transpose of B.

As U is positive semi-definite, we have

$$| U | \geq 0. \tag{3.1.11}$$

The above can further be reduced to

$$|W| \cdot |A - BW^{-1}B'| \geq 0, \tag{3.1.12}$$

which leads to

$$|A - B \cdot W^{-1} \cdot B'| \geq 0, \text{ as } W \text{ is positive definite.} \tag{3.1.13}$$

By the use of (3.1 8) we obtain from above

$$|A - C| > 0 \tag{3.1.14}$$

where C is the top left part of W^{-1}, consisting of s rows and s columns.

Let us now consider the matrix

$$||V_{ij} - v_{ij}||, \qquad (i, j = 1, 2, \cdots, T), \tag{3.1.15}$$

where $||v_{ij}||$ is the topleft part of W^{-1} consisting of T rows and T columns, and is equal to

$$||W_{11} - W_{12}W_{22}^{-1}W_{21}||^{-1}, \tag{3.1 16}$$

when W is written as

$$W = \begin{Vmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{Vmatrix}, \tag{3.1.17}$$

where W_{11} has T rows and T columns

By the repeated application of (3.1.14), we are led to the conclusion that all the leading minors of the matrix in (3.1.15) are either positive or zero. Hence the matrix in (3.1.15) is semi-positive definite.

If now we put

$$||\mu_{ij}|| = ||v_{ij}||^{-1} \tag{3.1.18}$$

we obtain

$$||\mu_{ij} - V^{ij}|| \tag{3.1.19}$$

to be semi-positive definite. Thus the ellipsoid

$$\sum_{i,j=1}^{T} V^{ij} \cdot t_i \cdot t_j = T + 2 \tag{3.1.20}$$

contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} \mu_{ij} \cdot t_i \cdot t_j = T + 2. \tag{3.1.21}$$

Cramer calls the ellipsoid in (3.1.20) a "concentration" ellipsoid.

We will now show that the ellipsoid given by (3.1.21) contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} I_{ij} \cdot t_i \cdot t_j = T + 2 \tag{3.1.22}$$

where $\| I_{ij} \|$ is the information matrix given by W_{11} in (3.1.17). We will prove the above by showing

$$\| I_{ij} - \mu_{ij} \|, \qquad (i, j = 1, \cdots, T), \tag{3.1.23}$$

to be semi-positive definite.

We obtain, from (3.1.16) and (3.1.18),

$$\| \mu_{ij} \| = W_{11} - W_{12}W_{22}^{-1}W_{21}, \qquad (i, j = 1, 2, \cdots, T). \tag{3.1.24}$$

From the above it follows that

$$\| I_{ij} - \mu_{ij} \| = W_{12}W_{22}^{-1}W_{21}. \tag{3.1.25}$$

Thus the matrix on the right hand side is semi-positive definite since W_{22} is positive definite, we see that the ellipsoid (3.1.21) contains within itself the ellipsoid given by (3.1.22). This proves the assertion made in (1.3.2) of §1.3. It may be seen that (3.1.22) is strictly contained in (3.1 21) if and only if $W_{12} \neq 0$. It may be mentioned that in this section as well as elsewhere, $T + 2$, appearing on the right hand side of the equation of an ellipsoid, can be replaced by any positive constant. Also the ellipsoid in (3.1 21) depends upon the choice of m and it can be shown that for any two positive integers m_1, m_2 $(m_1 > m_2)$ the ellipsoid for $m = m_1$ contains within itself the one for $m = m_2$.

§3.2. In general, let $\theta_i^*(x_1, x_2, \cdots, x_n)$ be statistics whose expectations are $\gamma_i(\theta_1, \theta_2, \cdots, \theta_T)$, $(i = 1, 2, \cdots, T)$, the latter being assumed to admit partial derivatives of all possible orders. Under the postulates enumerated in §3.1, we see that the ellipsoid in (3.1.20) contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} S_{ij} \cdot t_i \cdot t_j = T + 2 \tag{3.2.1}$$

where

$$\| S_{ij} \| = \| RW^{-1}R' \|^{-1}, \qquad i, j = 1, 2, \cdots, T, \tag{3.2.2}$$

and

$$R = \left\| \frac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} \gamma_j(\theta_1, \theta_2, \cdots, \theta_T) \right\|, \tag{3.2.3}$$

$$(j = 1, 2, \cdots, T; i_1 + i_2 + \cdots + i_T \leq m),$$

where j and $i_1 + i_2 + \cdots + i_T$ indicate the number of the row and the column respectively and is arranged to correspond to the arrangement of W, where W is the same as given in (3.1.9).

4. Achievement of the different lower bounds. In §4.1 we will demonstrate the desirability of finding a higher lower bound to the variance of sequential estimates than that given by Wolfowitz, by giving two examples in which the latter is not achieved From §2.4 it is clear that this will be so if $E(\phi_1(n) \cdot \phi_i(n))$ is not zero for at least one value of $i \geq 2$. We will demonstrate that this is true

for $i = 2$. In §4.2 we show that if "efficient" statistic exists for all $M \geq N$, the bound is achieved only in the case when the sample size is fixed. In §4.3 we obtain necessary and sufficient conditions for the attainment of the bound given in (1.1.4). In §4.4 we discuss the conditions under which there exists a "concentration ellipsoid" which coincides with the ellipsoid given in (3.1.21) for samples of fixed size N.

§4.1. Ex. 1. The Wald sequential procedure for testing $\theta = \theta_1$, against $\theta = \theta_2$ in a normal population

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \tag{4.1.1}$$

is given as follows: If

$$B < \sum_{i=1}^{s} \left(x_i - \frac{\theta_1 + \theta_2}{2} \right) < A, (s = 1, 2, \cdots, j-1), \tag{4.1.2}$$

and

$$\sum_{i=1}^{j} \left(x_i - \frac{\theta_1 + \theta_2}{2} \right) \text{ is either } \geq A \text{ or } \leq B, \tag{4.1.3}$$

we cease sampling and make a decision. Here A and B are constants fixed by the probability levels of making a correct decision

Let us denote the set of points satisfying (4.1.2) and (4.1.3) by R_j. In this case

$$\phi_1(n) = \sum_{i=1}^{n} (x_i - \theta) = Z_n - n\theta, \quad \text{where} \quad Z_n = \sum_{i=1}^{n} x_i. \tag{4.1.4}$$

The above is differentiable with respect to θ. On differentiating we have

$$\phi_2(n) = (Z_n - n\theta)^2 - n. \tag{4.1.5}$$

Now

$$E(\phi_1(n) \cdot \phi_2(n)) = E(Z_n - n\theta)^3 - E(n(Z_n - n\theta)). \tag{4.1.6}$$

By theorem 7.3, Wolfowitz [5],

$$E(Z_n - n\theta)^3 = En \cdot E(X - \theta)^3 + 3E(n(Z_n - n\theta)), \tag{4.1.7}$$

where X has the distribution given in (4.1.1). As $E(X - \theta)^3$ is equal to zero, (4.1.6) reduces to

$$E(\phi_1(n) \cdot \phi_2(n)) = 2E(n(Z_n - n\theta)). \tag{4.1.8}$$

We will now show that right hand side of (4.1.8) is not identically zero in θ. Let us consider

$$E(n) = \sum_{j=1}^{\infty} \int_{R_j} \frac{J}{(2\pi)^{j/2}} \cdot \left[\exp\left(-\tfrac{1}{2} \sum_{i=1}^{j} (x_i - \theta)^2 \right) \right] \prod_{u=1}^{j} dx_u. \tag{4.1.9}$$

Differentiating with respect to θ, we get

$$\frac{d}{d\theta}(E(n)) = \sum_{j=1}^{\infty} \int_{R_j} \frac{j(z_j - j\theta)}{(2\pi)^{j/2}} \cdot \left[\exp\left(-\tfrac{1}{2}\sum_{i=1}^{j}(x_i - \theta)^2\right)\right] \prod_{u=1}^{j} dx_u . \tag{4.1.10}$$

The righthand side of the above equation being equal to $E(n(Z_n - n\theta))$, the latter does not vanish identically in θ, because the lefthand side is not identically zero. The step from (4.1.9) to (4.1.10) can be easily seen to be valid.

Ex. 2. The Wald sequential procedure for testing $p = p_1$ against $p = p_2$ in a binomial distribution, where p is the probability of the event occurring, is given as follows: *If*

$$B < \sum_{i=1}^{s}(x_i - d) < A, \qquad s = 1, 2, \cdots, j-1, \tag{4.1.11}$$

and

$$\sum_{i=1}^{j}(x_i - d) \quad \textit{is either} \quad \geq A \quad \textit{or} \quad \leq B, \tag{4.1.12}$$

where d is given by $[\log(1-p_1)/(1-p_2)]/\log[(p_2(1-p_1)/p_1(1-p_2)]$, *the process stops with the jth observation and a decision is taken.* Here, x_i is the characteristic function of the event at the ith trial, that is:

$x_i = 1$, when the event occurs at the ith trial;

$= 0$, otherwise.

Let us denote the set of points satisfying (4.1.11) and (4.1.12) by R_j. In this case we find

$$E[\phi_1(n)\cdot\phi_2(n)] = \frac{2}{p^2\cdot(1-p)^2}\cdot E(n(Z_n - np)), \tag{4.1.13}$$

where $Z_n = \sum_{i=1}^{n} x_i$. We have now to show that the righthand side is not identically zero. Differentiating

$$E(n) = \sum_{j=1}^{\infty}\sum_{R_j} jp^{z_j}(1-p)^{j-z_j} \tag{4.1.14}$$

with regard to p, we obtain

$$\frac{d}{dp}(E(n)) = \sum_{j=1}^{\infty}\sum_{R_j}\frac{j(Z_j - jp)}{p(1-p)}\cdot p^{z_j}\cdot(1-p)^{j-z_j}. \tag{4.1.15}$$

The righthand side of the above is the same as

$$\frac{1}{p(1-p)}E(n(Z_n - np)). \tag{4.1.16}$$

Thus, the lefthand side of (4.1.15) being not identically zero, the same is true for (4.1.16), and consequently the bound given by Wolfowitz is not achieved in this case.

The step from (4.1.14) to (4.1.15) is valid as

$$\sum_{j=1}^{\infty}\left(\sum_{R_j} j\cdot p^{z_j}(1-p)^{j-z_j}\cdot\frac{z_j-j\cdot p}{p(1-p)}\right) \tag{4.1.17}$$

is absolutely and uniformly convergent.

§4.2. Let θ^* be some unbiased estimate of θ, where x_i's are successive independent observations on the chance variable X having the probability density function or probability function $f(x;\theta)$. We adopt a sequential procedure mentioned in §2.1 satisfying the regularity conditions in §2.2 and also postulate the following:

(i) For all positive integral values of $M \geq N$

$$p_M(x_1, x_2, \cdots, x_M; \theta) = \prod_{i=1}^{M} f(x_i; \theta)$$

possesses an 'efficient' estimate for θ, where N is the least value of n for which $\Pr(n = N) \neq 0$.

(ii) $E(n)$ exists and admits derivatives up to the second order with respect to θ. Furthermore, $\frac{d}{d\theta}(E(n))$ is either zero for all θ under consideration or is never zero.

Under the above conditions the Wolfowitz lower bound for the variance of unbiased estimates is achieved only when $\Pr(n = N) = 1$.

PROOF: This bound will be attained if and only if there exists an unbiased estimate θ^* of θ such that

$$E(\theta^* - \theta - K\phi_1(n))^2 = 0, \tag{4.2.1}$$

that is,

$$\theta^* - \theta = K\phi_1(n) \tag{4.2.2}$$

with probability one, where K is independent of all x_i's and n. As there exists an 'efficient' estimate, say $\psi(M)$ for all $M \geq N$, we have

$$\psi(M) - \theta = \frac{1}{M\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]} \tag{4.2.3}$$

for all $M \geq N$. From (4.2.2) and (4.2.3), it follows that

$$\theta^* - \theta = K\cdot n\cdot(\psi(n) - \theta)\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]. \tag{4.2.4}$$

Now as

$$K = \frac{1}{En\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]}, \tag{4.2.5}$$

we have

$$\theta^* - \theta = \frac{n \cdot (\psi(n) - \theta)}{En}. \tag{4.2.6}$$

If $E(n)$ is independent of θ, then from (4.2.6), we obtain

$$n/E(n) = 1, \tag{4.2.7}$$

that is, n is constant with probability one and the sequential procedure reduces to a fixed size sample case. If $E(n)$ is not independent of θ, then differentiating (4.2.6) with regard to θ, we obtain

$$1 = \frac{n \cdot (\psi(n) - \theta) \cdot \frac{d}{d\theta}(En)}{(En)^2} + \frac{n}{En}. \tag{4.2.8}$$

As $\frac{d}{d\theta}(En)$ is not equal to zero for any θ under consideration, substituting the value of $\psi(n)$ from (4.2.8) in (4.2.6), the latter takes the form:

$$\theta^* - \theta = \frac{En - n}{\frac{d}{d\theta}(En)}. \tag{4.2.9}$$

Differentiating the above with respect to θ, the result is:

$$-1 = -\frac{En - n}{\left[\frac{d}{d\theta}(En)\right]^2} \cdot \frac{d^2}{d\theta^2}(En) + 1. \tag{4.2.10}$$

Now if $\frac{d^2}{d\theta^2}(En) = 0$, then (4.2.10) is not valid, thereby contradicting (4.2.2). If $\frac{d^2}{d\theta^2}(En) \neq 0$, then rearranging (4.2.10), we obtain

$$n = -\frac{2\left(\frac{d}{d\theta} En\right)^2}{\frac{d^2}{d\theta^2}(En)} + En, \tag{4.2.11}$$

that is, n is a constant with probability one. This proves that Wolfowitz bound is achieved only in the case when $n = N$ with probability one. This generalizes the result of Blackwell and Girshick [6][2] to the extent that in [6] the existence of an efficient estimate is assumed for all integral values of M instead of $M \geq N$, as assumed here. Moreover the proof given here, with slight modifications, is also valid when the successive observations are not independent.

[2] In [6] the assumption that "$x_1 + x_2 + \cdots + x_M$ be a sufficient statistic for all M" really amounts to the postulate that "$x_1 + x_2 + \cdots + x_M$ be an "efficient" statistic for all M," when we restrict ourselves to probability density functions satisfying the conditions given by Koopman in [7].

§4.3. Let us consider a sample of fixed size N. Let θ^* together with the probability density function p_N satisfy the following regularity conditions:

(i). There exists a transformation T from $(x_1, x_2, \cdots, x_N)$ to the variables

$$\xi_i = \xi_i(x_1, x_2, \cdots, x_N), \qquad \theta^* = \theta^*(x_1, x_2, \cdots, x_N), \tag{4.3.1}$$
$$i = 1, 2, \cdots, N-1,$$

such that

(a). The functions ξ_i are everywhere unique and continuous, and have continuous partial derivatives

$$\frac{\partial \xi_i}{\partial x_u}, \frac{\partial \theta^*}{\partial x_u} \quad (i = 1, 2, \cdots, N-1, u = 1, 2, \cdots, N)$$

in all points $(x_1, x_2, \cdots, x_N)$ except possibly in certain points belonging to a finite number of hyper-surfaces.

(b). The relation (4.3.1) define a one-to-one correspondence between the points $x = (x_1, x_2, \cdots, x_N)$ and $y = (\xi_1, \xi_2, \cdots, \xi_{N-1}, \theta^*)$ so that conversely $x_i = \eta_i(\xi_1, \xi_2, \cdots, \xi_{N-1}, \theta^*)$ where η_i are unique.

(ii). There exists partial derivatives of $g(\theta^*;\theta)$, $h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta)$ with regard to θ of all orders up to and including m, where m is some finite integer. The variances of θ^*, h_i and $g_i \cdot g_j$, $i, j = 1, 2, \cdots, m$, are finite, where h_i and g_i are defined in section 1.

(iii). There exist functions

$$T_{ij} \begin{pmatrix} i = 1, 2, \cdots, m \\ j = 1, 2, 3 \end{pmatrix}$$

such that

$$\left| \frac{\partial^i p_N}{\partial \theta^i} \right| < T_{i1}(x_1, x_2, \cdots, x_N);$$

$$\left| \frac{\partial^i g}{\partial \theta^i} \right| < T_{i2}(\theta^*);$$

$$\left| \frac{\partial^i h}{\partial \theta^i} \right| < T_{i3}(\xi_1, \xi_2, \cdots, \xi_{N-1}; \theta^*),$$

for all θ in D and for almost all $(x_1, x_2, \cdots, x_N)$ where D is an open interval. Further

$$\int T_{i1}(x_1, x_2, \cdots, x_N) \prod_{u=1}^{N} dx_u,$$

$$\int T_{i2}(\theta^*)\, d\theta^* \quad \text{and} \quad \int T_{i3}(\xi_1, \xi_2, \cdots, \xi_{N-1}; \theta^*) \prod_{i=1}^{N-1} d\xi_u$$

are all finite, the range of integration, in each case, being the whole range for the arguments indicated. Then the necessary and sufficient conditions that the variance of θ^* equals the lower bound given in (1.1.4) are

(iv). $h_1, h_2, \cdots, h_m$ are linearly dependent considered as functions of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for any given θ^* and θ, and

(v) The probability density function g of θ^* is of the form

$$\theta^* - \theta = \sum_{i=1}^{m} K_i g_i$$

where K_i may depend upon θ and N only.

The proof here is given when p_N is a probability density function. It is also valid with slight modification when p_N is the probability of discrete variables.

PROOF: Let J be the Jacobian of the transformation T in (4.3.1). Then because of conditions (i) and (ii) above, we have,

$$p_N(x_1, x_2, \cdots, x_N; \theta) \cdot |J| = g(\theta^*; \theta) \cdot h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta) \tag{4.3.2}$$

Further

$$\int h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta) \prod_{u=1}^{N-1} d\xi_u = 1, \tag{4.3.3}$$

the range of integration being the space of $\xi_1, \xi_2, \cdots, \xi_{N-1}$. Differentiating the above i times under the integral sign, it follows that

$$E(h_i \mid \theta^*, \theta) = 0. \tag{4.3.4}$$

Similarly we have

$$E(g_i \cdot h_i) = 0 \tag{4.3.5}$$

as the expectation of the quantity on the L.H.S. is finite by virtue of (ii). More generally, we have

$$E(F(\theta^*) \cdot h_i) = E[F(\theta^*) \cdot E(h_i \mid \theta^*)] = 0 \tag{4.3.6}$$

if $E(F(\theta^*) \cdot h_i)$ is finite. Let us now examine

$$E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i \phi_i(N)\right)^2, \tag{4.3.7}$$

where $K_i\phi_i(N)$ can also be written as

$$K_i\left(g_i + \binom{i}{1} h_1 g_{i-1} + \cdots + h_i\right). \tag{4.3.8}$$

Now (4.3.7) can be put in the form

$$E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i g_i - \sum_{i=1}^{m} L_i \cdot h_i\right)^2, \tag{4.3.9}$$

where

$$L_i = \sum_{j=1}^{m} K_j \cdot \binom{j}{i} \cdot g_{j-i}, \qquad (i = 1, 2, \cdots, m), \tag{4.3.10}$$

clearly depend on θ and θ^* only.

By virtue of (4.3.4–4.3.6) and $F(\theta^*)$ involved in (4.3.9) being such that $E[F(\theta^*) \cdot h_i](i = 1, 2, \cdots, m)$ is finite because of (ii), we can further reduce (4.3.9) to

$$E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i g_i\right)^2 + E\left[E\left(\left(\sum_{i=1}^{m} L_i h_i\right)^2 \,\middle|\, \theta^*\right)\right]. \tag{4.3.11}$$

The lower bound will be achieved if and only if the above expression is zero, the necessary and sufficient conditions for which are:

$$\theta^* - \theta = \sum_{i=1}^{m} K_i \cdot g_i , \tag{4.3.12}$$

and

$$\sum_{i=1}^{m} L_i h_i \equiv 0 \quad \text{in} \quad \xi_1 , \xi_2 , \cdots , \xi_{N-1} \tag{4.3.13}$$

for any given values of θ^* and θ.

(4.3.13) is equivalent to the condition that h_i, $(i = 1, 2, \cdots, m)$ are linearly dependent considered as functions of $\xi_1 , \xi_2 , \cdots , \xi_{N-1}$ for any given values of θ and θ^*.

When m takes the value one, the above reduces to the Cramer conditions for the existence of an "efficient" estimate.

§4.4. *Multiparameter case.* Let $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ be the unbiased estimates of $\theta_1 , \theta_2 , \cdots , \theta_T$ in the probability density function

$$p_N(x_1 , x_2 , \cdots , x_N ; \theta_1 , \theta_2 , \cdots , \theta_T)$$

and the regularity conditions of §4.3 are satisfied when θ^* and $\frac{\partial^i}{\partial \theta^i}$ $(i = 1, 2, \cdots, m)$ are replaced by θ_j^* $(j = 1, 2, \cdots, T)$ and

$$\frac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} \qquad i_1 + i_2 + \cdots + i_T \leq m)$$

respectively. Further let

$$\begin{aligned} &p_N(x_1 , x_2 , \cdots , x_N ; \theta_1 , \theta_2 , \cdots , \theta_T) \cdot |J| \\ &\quad = g(\theta_1^*, \theta_2^*, \cdots , \theta_T^*; \theta_1 , \theta_2 , \cdots , \theta_T) \\ &\qquad \cdot h(\xi_1 , \xi_2 , \cdots , \xi_{N-1} \mid \theta_1^*, \theta_2^*, \cdots , \theta_T^*) \end{aligned} \tag{4.4.1}$$

where g and h are respectively the joint probability distribution functions of $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ and the conditional probability distribution of $\xi_1 , \xi_2 , \cdots , \xi_{N-1}$ for a given set of values of $\theta_1^*, \theta_2^*, \cdots , \theta_T^*$. In order that the ellipsoid (3.1.20) coincides with the one given by (3.1.21), it is necessary and sufficient that the following be satisfied for each $t(t = 1, 2, \cdots, T)$

$$E\left(\theta_t^* - \theta_t - \sum_{i_1+i_2+\cdots+i_T \leq m} {}^{(t)}K_{i_1,i_2,\cdots,i_T} \cdot \phi_{i_1,i_2,\cdots,i_T}^{(N)}\right)^2 = 0. \tag{4.4.2}$$

Now reasoning similar to that in §4.3, we conclude from the above that the necessary and sufficient conditions are.

There exist T independent linear combinations of

$$h_{r_1 r_2 \cdots r_T}, \quad r_1 + r_2 + \cdots + r_T \leq m \tag{4.4.3}$$

which vanish with probability one for any given values of the sets

$$(\theta_1^*, \theta_2^*, \cdots, \theta_T^*) \quad \text{and} \quad (\theta_1, \theta_2, \cdots, \theta_T),$$

and

$$\theta_i^* - \theta_i = \sum_{r_1+r_2+\cdots+r_T \leq m} {}^{(i)}K_{r_1 r_2 \cdots r_T} \theta_1^{r_1} \theta_2^{r_2} \cdots \theta_T^{r_T}, \qquad i = 1, 2, \cdots, T, \tag{4.4.4}$$

where the K's do not depend upon θ_i^* and ξ_i's For $T = 1$, the above reduce to the conditions in §4.3 We will now give an example in which (4.4.3) and (4.4.4) are satisfied Let

$$p_N(x_1, x_2, \cdots, x_N, \theta_1, \theta_2) = \frac{1}{(2\pi\theta_1)^{N/2}} \left[\exp - \frac{1}{2\theta_1} \cdot \sum_{i=1}^{N} (x_i - \theta_2)^2 \right] \tag{4.4.5}$$

We have

$$\theta_1^* = \sum_{i=1}^{N} (x_i - \bar{x})^2/(N-1), \tag{4.4.6}$$

$$\theta_2^* = \sum_{i=1}^{N} x_i/N = \bar{x}, \tag{4.4.7}$$

unbiased estimates of θ_1 and θ_2 in (4.4.5) The joint distribution of θ_1^* and θ_2^* is given by

$$g(\theta_1^*, \theta_2^*, \theta_1, \theta_2) = C \exp \left[\frac{-N(\theta_2^* - \theta_2)^2 - (N-1)\theta_1^*}{2\theta_1} \right] (\theta_1^*)^{N-1/2} \cdot (\theta_1)^{-N/2}, \tag{4.4.8}$$

It can be easily seen that the condition (4.4.3) is satisfied, and the estimates themselves can be put in the form

$$\theta_1^* = \theta_1 + \frac{2\theta_1^2}{N-1} \cdot \frac{1}{g} \frac{\partial g}{\partial \theta_1} - \frac{\theta_1^2}{N(N-1)} \cdot \frac{1}{g} \frac{\partial^2 g}{\partial \theta_1^2}, \tag{4.4.9}$$

$$\theta_2^* = \theta_2 + \frac{\theta_1}{N} \frac{1}{g} \frac{\partial g}{\partial \theta_2}. \tag{4.4.10}$$

It is thus seen that the 'concentration' ellipsoid for θ_1^*, θ_2^* coincides with the ellipsoid (3.1.21) for $m = 2$. On the other hand if we use $m = 1$, the condition (4.4.3) is satisfied but not the one in (4.4.4), as can be seen from (4.4.9), and thus the concentration ellipsoid strictly contains within itself the one given by the information matrix It may be noted that for $m = 1$, the condition (4.4.3)

merely requires that a system of sufficient statistics exists for estimating θ_1, θ_2, $\cdots$, θ_T. The reason is that the condition (4.4.3) takes the equivalent form

$$\frac{\partial h}{\partial \theta_\iota} = 0 \tag{4.4.11}$$

for $\iota = 1, 2, \cdots, T$ identically in $\xi_1, \xi_2, \cdots, \xi_{N-T}$ that is, that h is free of $\theta_1, \theta_2, \cdots, \theta_T$.

5 Miscellaneous In §5.1–§5.3 we discuss certain properties of $\phi_\iota(n)$. In §5.4 we obtain conditions under which there exists no unbiased estimate of θ, having a finite variance, which is functionally dependent upon a given unbiased estimate θ^* of θ.

§5.1 Assume that there exists an "efficient" statistic $\theta^*(x_1, x_2, \cdots, x_N)$ for estimating θ, in probability density function (or probability)

$$p_N(x_1, x_2, \cdots, x_N, \theta)$$

That is,

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = K \quad \phi_1(N) \tag{5.1.1}$$

where K as usual may only depend on θ. We postulate as usual the existence of all partial derivatives of p_N of all orders and also of K up to the third order with

$$\frac{d^3 K}{d\theta^3} = 0. \tag{5.1.2}$$

Further we assume that

$$\left| \frac{\partial^\iota p_N}{\partial \theta^\iota} \right| < T_\iota(x_1, x_2, \cdots, x_N)$$

where

$$\int T_\iota(x_1, x_2, \cdots, x_N) \prod_{u=1}^{N} dx_u \qquad \text{is finite for all } \iota$$

Under the above assumptions we will show that

$$\phi_0(N) = 1, \phi_1(N), \phi_2(N), \cdots, \phi_\iota(N), \cdots$$

form a set of orthogonal polynomials in $\phi_1(N)$ with respect to the weight function

$$p_N(x_1, x_2, \cdots, x_N; \theta).$$

Proof. We can easily see that

$$\frac{\partial \phi_\iota}{\partial \theta} = \phi_{\iota+1} - \phi_\iota \, \phi_1 \tag{5.1.3}$$

where $\phi_\iota(N)$ is shortened to ϕ_ι for convenience. Differentiating (5.1.1) with respect to θ,

$$\frac{\partial \phi_1}{\partial \theta} = -\frac{1}{K} \frac{dK}{d\theta} \phi_1 - \frac{1}{K} \tag{5.1.4}$$

Let us designate

$$z_i = \frac{1}{K}\frac{d^i K}{d\theta^i} \tag{5 1 5}$$

for all integral values of i. From (5 1 3) and (5 1 4), it follows that

$$\phi_2 - \phi_1^2 = -z_1\phi_1 - \frac{1}{K}, \tag{5 1.6}$$

Differentiating (5.1.6) further with regard to θ and using (5 1 3) and (5 1 6) we obtain

$$\phi_3 - \phi_1\phi_2 = -2z_1\phi_2 - \left(\frac{2}{K} + z_2\right)\phi_1. \tag{5.1 7}$$

Differentiating (5 1.7) with regard to θ, and using (5 1 2) we get

$$\phi_4 - \phi_1\phi_3 = -3z_1\phi_3 - \left(3z_2 + \frac{3}{K}\right)\phi_2. \tag{5.1.8}$$

We assume generally that

$$\phi_{i+1} - \phi_1\phi_i = -iz_1\phi_i - \left(\frac{i(i-1)}{2}z_2 + \frac{i}{K}\right)\phi_{i-1}. \tag{5 1 9}$$

Differentiating (5 1 9), and employing (5 1 3), (5 1 3) and (5.1.9) we obtain

$$\phi_{i+2} - \phi_1\phi_{i+1} = -(i+1)z_1\phi_{i+1} - \left(\frac{i(i+1)}{2}z_2 + \frac{i+1}{K}\right)\phi_i. \tag{5.1 10}$$

We know that (5.1.9) holds for $i = 1, 2, 3$, ϕ_0 being taken equal to one, and we have proved that if (5 1 9) is true for $i = j$, it is true for $i = j + 1$. Thus by mathematical induction (5 1.9) holds good for all integral values of i.

It is also clear from (5 1 6) and (5 1.9) that ϕ_i can be expressed as a polynomial in ϕ_1 of the ith degree, the coefficient of ϕ_1^i being equal to unity.

To complete the proof of our assertion we will now prove that

$$E(\phi_i \cdot \phi_j) = 0, \qquad i \neq j \tag{5.1 11}$$

From (5.1 9)

$$\phi_1\ \phi_i = \phi_{i+1} + iz_1\phi_i + \left(\frac{i(i-1)}{2}z_2 + \frac{i}{K}\right)\phi_{i-1}, \tag{5.1 12}$$

where i is any positive integer. We multiply both sides of (5 1 12) by ϕ_1 and reduce every product $\phi_i\phi_1$ to a linear combination of ϕ_{j+1}, ϕ_j and ϕ_{j-1} with the help of (5 1 12) Repeating this process $j - 1$ times ($j < i$) it follows that:

$$\phi_1^j \cdot \phi_i = \phi_{i+j} + \sum_{u=1}^{2j-1} d_u^j\ \phi_{i+j-u} + d_{2j}^j \cdot \phi_{i-j} \tag{5 1 13}$$

where d_u^j are functions of K, z_1 and z_2. From (5 1.13), by taking expectations of both sides,

$$E(\phi_1^j\ \ \phi_i) = 0, \qquad (j < i). \tag{5 1.14}$$

Now, since ϕ_i is a polynomial of the jth degree in ϕ_1 we conclude that (5 1 11) is true for all integral (positive) values of i

Thus we obtain

(5 1 15) $\phi_0(N) = 1, \phi_1(N), \phi_2(N), \cdots, \phi_i(N), \cdots$

as a set of orthogonal polynomials in $\phi_1(N)$, the weight function being

$$p_N(x_1, x_2, \cdots, x_N, \theta)$$

Furthermore

$$\phi_1^{i-1}\,\phi_i = \phi_{2i-1} + \sum_{u=1}^{2i-3} d_u^{i-1}\,\phi_{2i+1-u} + d_{2i-2}^{i-1}\,\phi_i \tag{5 1 16}$$

where

$$d_{2i-2}^{i-1} = \prod_{j=2}^{i} B_j \tag{5 1.17}$$

and

$$B_j = \frac{j(j-1)}{2}\,z_2 + \frac{j}{K}. \tag{5 1 18}$$

Hence

$$E(\phi_1^i\,\phi_i) = \prod_{j=1}^{i} B_j \tag{5 1 19}$$

Thus if we divide ϕ_i by $\sqrt{\prod_{j=1}^{i} B_j}$, (5 1.15) becomes the orthonormal set.

Some cases, where we obtain ϕ_i as orthogonal polynomials, are given below,

1 $p_N = \dfrac{1}{(\sqrt{2\pi})^N}\; e^{-\frac{1}{2}\sum_1^N (x_i-\theta)^2}, \qquad \phi_1 = \sum_{i=1}^{N} (x_i - \theta).$

2 $p_N = \dfrac{1}{(2\pi\theta)^{N/2}}\; e^{-\frac{1}{2\theta}\sum_1^N x_i^2}, \qquad \phi_1 = \dfrac{\sum_{i=1}^{N} x_i^2}{2\theta^2} - \dfrac{N}{2\theta}.$

3. $p_N = \theta^{\sum_1^N x_i} \cdot (1-\theta)^{N-\sum_1^N x_i}$ $\begin{pmatrix} x_i = 1 \text{ with prob. } \theta \\ = 0 \text{ with prob. } 1-\theta \end{pmatrix}$,

$$\phi_1 = \frac{\sum_{i=1}^{N} x_i - N\theta}{\theta(1-\theta)}.$$

4 $p_N = \dfrac{e^{-N\theta}\cdot\theta^{\sum_1^N x_i}}{\prod_{i=1}^{N} x_i!}, \qquad \phi_1 = \dfrac{\sum_{i=1}^{N} x_i - N\theta}{\theta}.$

A_i and B_i, the coefficients of ϕ_i and ϕ_{i-1} respectively in (5.1.12) for the above four cases are given as below:

	A_i	B_i
1	0	$i\ N$
2	$2i/\theta$	$\frac{i(i-1)}{\theta^2} + \frac{iN}{2\theta^2}$
3	$\frac{i(1-2\theta)}{\theta(1-\theta)}$	$\frac{-i(i-1)}{\theta(1-\theta)} + \frac{iN}{\theta(1-\theta)}$
4	i/θ	iN/θ

It may be mentioned that in all these cases $\{\phi_i\}$ are also a complete set of polynomials

§5 2. Let $\sum_{i=1}^{m} K_i\phi_i(n)$, where $K_i (i = 1, 2, \cdots, m)$ depends upon θ be such that $\sum_{i=1}^{m} K_i\phi_i(n)$ and $\phi_i(n)$ satisfy the regularity conditions mentioned in §2 2 Then we will show that $\sum_{i=1}^{m} K_i\phi_i(n)$ cannot be a function of $x_1, x_2, \cdots, x_n$ alone except for constant zero

PROOF: Let us assume that $\sum_{i=1}^{m} K_i \cdot \phi_i(n)$ is independent of θ, that is, it is some statistic, say,

$$\theta^*(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{m} K_i \cdot \phi_i(n). \tag{5.2 1}$$

Taking expectations of both the sides, we obtain

$$E(\theta^*(x_1, x_2, \cdots, x_n)) = \sum_{i=1}^{m} K_i\ E\phi_i(n) = 0. \tag{5 2 2}$$

Differentiating (5 2 2) i times with regard to θ, we have, because of the regularity conditions on $\phi_i(n)$ and $\theta^*(x_1, x_2, \cdots, x_n)$,

$$E(\theta^*(x_1, x_2, \cdots, x_n)\ \ \phi_i(n)] = 0, \qquad i = 1, 2, \cdots, m. \tag{5 2 3}$$

It may be noted this is similar to the result in (2 3) From (5 2.3) and (5 2.1) it follows that

$$E[\theta^*(x_1, x_2, \cdots, x_n)]^2 = 0 \tag{5.2 4}$$

Thus $\theta^*(x_1, x_2, \cdots, x_n)$ is zero with probability one, that is,

$$\sum_{i=1}^{m} K_i\ \phi_i(n),$$

if independent of θ, is zero with probability one This proves our assertion that this cannot be a function of $x_1, x_2, \cdots, x_n$ alone except for constant zero

From the foregoing we deduce the following conclusions

I $\phi_i(n)$ or any power of it cannot be a function of the observations free of θ.

II If a statistic $\theta^*(x_1, x_2, \quad, x_n)$, which is not a constant with probability one, can be put in the form

$$\theta^*(x_1, x_2, \cdot \quad, x_n) = K_0 + \sum_{i=1}^{m} K_i \phi_i(n), \tag{5 2 5}$$

where m is some finite positive integer, then

(i) K_0 must depend upon θ,

(ii) The expression (5 2 5) for $\theta^*(x_1, x_2, \cdot \quad, x_n)$ in $\phi_i(n)$ is unique.

(iii) No other unbiased estimate of K_0 satisfying the regularity conditions can be put in the form (5 2.5)

(iv) When $m = 1$, there is no other statistic except $a\theta^* + b$, where a and b are constants independent of θ, which can be put in the above form $K_0 + K_1 \quad \phi_1(n)$, K_0 and K_1 are differentiable functions of θ and K_1 does not vanish for any θ under consideration

(v) Let ξ be any function of $x_1, x_2, \quad, x_n$ free of θ, satisfying the regularity conditions of §2 2 with $E(\xi) = 0$ Since the covariance between ξ and $\theta^*(x_1, x_2, \cdot \quad, x_n)$ in (5 2 5) is equal to zero, the statistic of the form (5 2 5) has the least variance of all unbiased estimates of K_0 that satisfy the regularity conditions of §2 2

Also, if the probability density or the probability function depends on more than one parameter, then all the above results except (iv) hold good if

$$\sum_{i=1}^{m} K_i \phi_i (n)$$

is replaced by

$$\sum_{i_1+i_2+\cdots+i_r \le m} K_{i_1,i_2, \quad ,i_r} \phi_{i_1,i_2, \quad ,i_r}(n).$$

§5 3 Let us now prove the assertion made in (iv) of §5 2, when m is equal to one

Suppose the contrary that there is a statistic $\theta_1^*(x_1 \quad x_2, \cdots \quad, x_n)$ which is of the form

$$\theta_1^*(x_1, x_2, \quad, x_n) = L_0 + L_1 \cdot \phi_1(n), \tag{5 3 1}$$

$\theta^*(x_1, x_2, \quad, x_n)$, of course, has the form

$$\theta^*(x_1, x_2, \quad, x_n) = K_0 + K_1 \cdot \phi_1(n), \tag{5 3 2}$$

We will assume K_0, K_1, L_0, L_1 to be differentiable functions of θ and that K_1, L_1 do not vanish for values of θ under consideration

Differentiating, with respect to θ, the expressions in (5 3 1) and (5.3.2, we have

$$\frac{dL_0}{d\theta} + \frac{dL_1}{d\theta} \cdot \phi_1 + L_1(\phi_2 - \phi_1^2) = 0, \tag{5 3 3}$$

$$\frac{dK_0}{d\theta} + \frac{dK_1}{d\theta} \quad \phi_1 + K_1(\phi_2 - \phi_1^2) = 0, \tag{5 3 4}$$

where ϕ_i is short for $\phi_i(n)$ Taking the expectations of the above and rearranging, it follows that

$$E(\phi_1^2) = \frac{1}{L_1}\frac{dL_0}{d\theta} = \frac{1}{K_1}\frac{dK_0}{d\theta}. \tag{5 3.5}$$

From (5 3 3) to (5 3 5), we deduce that

$$\frac{1}{L_1}\frac{dL_1}{d\theta} = \frac{1}{K_1}\frac{dK_1}{d\theta}. \tag{5 3 6}$$

Now solving the above differential equation, we get

$$L_1 = aK_1, \tag{5 3 7}$$

where a is a constant independent of θ. From (5 3 5) and (5 3.7) it follows that

$$L_0 = aK_0 + b, \tag{5.3 8}$$

where b is a constant independent of θ From (5.3 7) and (5 3 8) we conclude that the statistic in (5 3 1) must be of the form $a\theta^* + b$, which proves our assertion An immediate consequence is that if there exists an efficient statistic for estimating $\gamma(\theta)$, then no other function of θ except a $\gamma(\theta) + b$ can have an efficient estimate.[3]

§5 4. If $\theta^*(x_1, x_2, \cdots, x_n)$ is an unbiased estimate of θ satisfying the following conditions

(i) Among all unbiased estimates of θ having finite variances, which are also functions of θ^*, θ^* is one with the least variance,

(ii) For all θ there exists a complete set of polynomials with respect to the distribution function of θ^*, then there exists no unbiased estimate of θ with a variance, which is functionally dependent upon θ^*, except θ^* itself.

PROOF. Let θ^* be the unbiased estimate of θ which has the least variance among all unbiased estimates of θ which are functions of θ^* Further let $S(\theta^*)$ be any function of θ^*, free of θ, whose expectation exists and is equal to zero Let the variance of $S(\theta^*)$ be finite It is well known that for any such $S(\theta^*)$

$$E(\theta^* S(\theta^*)) = 0 \tag{5 4.1}$$

Now $\theta^* S(\theta^*)$ in turn having expectation equal to zero, we obtain

$$E(\theta^{*2} S(\theta^*)) = 0. \tag{5 4 2}$$

Repeating the above i times we obtain, in general, that

$$E(\theta^{*i} S(\theta^*)) = 0 \tag{5 4 3}$$

[3] We assume the existence of $\frac{d^i\gamma}{d\theta^i}$ $(i = 1, 2)$ and $\frac{d}{d\theta}(E\phi_1^4)$ for all θ, and also postulate that $\frac{d\gamma(\theta)}{d\theta}$ and $E(\phi_1^4)$ do not vanish for any θ under consideration

for all positive integers i. From the above, with the help of condition (ii), we conclude that $S(\theta^*)$ must be equal to zero. Thus if $H(\theta^*)$ is an unbiased estimate of θ with finite variance, then from above, $H(\theta^*) - \theta^*$, having the expectation zero and a finite variance, must be zero with probability one. Thus $H(\theta^*)$ is the same as θ^*, which proves the result.

EXAMPLE. If θ^* is of the form (5.2.7) and condition (ii) is satisfied, then there is no function of θ^*, free of θ and having a finite variance, whose expectation is K_0.

Conditions (i) and (ii) above are satisfied for estimating θ in the examples quoted at the end of the section 5.1, and thus in these cases the result holds good when θ^* is the efficient estimate.

I am highly thankful to Professor J. Wolfowitz for his guidance and help in this research.

REFERENCES

[1] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946, p. 480.

[2] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Calcutta Math. Soc. Bull.*, September, 1945.

[3] A. BHATTACHARYYA, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946), also "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1947).

[4] H. CRAMÉR, "Contributions to the theory of statistical estimation," *Skandinavisk Aktuar tids.*, Vol. 29 (1946), pp. 5-94.

[5] J. WOLFOWITZ, "Efficiency of sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947).

[6] BLACKWELL AND GIRSHICK, "A lower bound for the variance of some unbiased sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947).

[7] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Am. Math. Soc. Trans.*, Vol. 39 (1936), p. 399.

ON THE THEORY OF SOME NON-PARAMETRIC HYPOTHESES

By E. L. Lehmann and C. Stein

University of California, Berkeley

Summary. For two types of non-parametric hypotheses optimum tests are derived against certain classes of alternatives. The two kinds of hypotheses are related and may be illustrated by the following example. (1) The joint distribution of the variables $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ is invariant under all permutations of the variables, (2) the variables are independently and identically distributed. It is shown that the theory of optimum tests for hypotheses of the first kind is the same as that of optimum similar tests for hypotheses of the second kind. Most powerful tests are obtained against arbitrary simple alternatives, and in a number of important cases most stringent tests are derived against certain composite alternatives. For the example (1), if the distributions are restricted to probability densities, Pitman's test based on $\bar{y} - \bar{x}$ is most powerful against the alternatives that the X's and Y's are independently normally distributed with common variance, and that $E(X_i) = \xi$, $E(Y_j) = \eta$ where $\eta > \xi$. If $\eta - \xi$ may be positive or negative the test based on $|\bar{y} - \bar{x}|$ is most stringent. The definitions are sufficiently general that the theory applies to both continuous and discrete problems, and that tied observations present no difficulties. It is shown that continuous and discrete problems may be combined. Pitman's test, for example, when applied to certain discrete problems, coincides with Fisher's exact test, and when $m = n$ the test based on $|\bar{y} - \bar{x}|$ is most stringent for hypothesis (1) against a broad class of alternatives which includes both discrete and absolutely continuous distributions.

1. Generalities. In the present paper we study the problem of determining optimum tests for certain non-parametric hypotheses. It is important in this connection to make some distinctions which are of lesser significance when the problem is approached from the intuitive point of view which has been customary in this field. Consider for example the hypothesis H that $Z_1, \cdots, Z_N$ are independently and identically distributed according to an unknown probability density function. All tests which have been suggested for testing H are valid also for testing the hypothesis H' that the unknown joint probability density function of the Z's is symmetric in its N arguments. On the other hand, tests which have optimum properties for testing H' against a certain class of alternatives will in general not possess the same properties when H' is replaced by H. From the present point of view the two hypotheses mentioned are essentially different. We shall be concerned in this paper primarily with generalizations of H', and we shall show that many of the tests suggested in the literature have optimum properties for testing hypotheses of this kind against certain classes of alternatives.

The corresponding general theory for hypotheses related to H is quite different.

However the two theorems do coincide, provided tests of these latter hypotheses are restricted to similar regions. More specifically, all results on optimum tests of H' are equivalent to the corresponding results on optimum similar tests of H, and this equivalence holds also for many of the more general hypotheses considered in this paper.

It should be observed that in many experimental situations, the hypothesis H' that the joint distribution of the Z's is invariant under all permutations is more realistic than the hypothesis H that the Z's are independently and identically distributed. For example, suppose there is a block of land divided into $m + n$ plots, and the experimenter wants to test whether one of two fertilizers (used in fixed amounts) is more effective than the other in increasing the yield of a certain plant. Of the plots, m are chosen at random, fertilizer I is applied to these, and fertilizer II to the other n. If X_i denotes the yield from the ith plot to which fertilizer I has been applied and Y_j denotes the yield from the jth plot to which fertilizer II has been applied, where the plots are numbered at random, then the hypothesis that the two fertilizers are completely equivalent implies that the application of any permutation to $X_1, \cdots, X_m, Y_1, \cdots Y_n$ does not change their joint distribution. But it is not reasonable to suppose the X_i, Y_j are independently and identically distributed, since there may be intrinsic differences among the plots. For discussions of these and related points, see Fisher [1], Neyman [2], Pitman [3]. It may be that in many particular cases some hypothesis between the two is really appropriate but the hypothesis H is the only one that is evidently appropriate from a cursory inspection of the setup.

Many of the alternative hypotheses considered below, for example those involving normality, are dictated more by tradition and ease of treatment than by appropriateness in actual experiments. Thus this paper should not be considered as providing absolute justification for tests such as Pitman's but rather as suggesting a method of obtaining optimum non-parametric tests when the class of alternatives is fairly well specified.

Another possibility, first raised by Neyman [2], which has been ignored in this paper is the equality on the average of the two fertilizers but with fertilizer I having a larger dispersion than fertilizer II, or a distribution differing in some other characteristic. It would be reasonable to consider this as part of the hypothesis tested, but tests based on randomization may give a probability of rejection of the hypothesis of equivalence in this case which is much higher than the stated level of significance. We hope to return to problems of this type in later papers.

Let us make the following basic assumptions. $\mathcal{Z}$ is a space of points z and $\mathcal{A}$ is an additive class of subsets A of $\mathcal{Z}$. Any member of $\mathcal{A}$ will be said to be measurable. By a probability distribution we mean a measure F, defined over $\mathcal{A}$ for which $F(\mathcal{Z}) = 1$. We shall be concerned with two classes of probability distributions. One, the class of all distributions, and two, the class of distributions which are absolutely continuous with respect to a given measure μ, that is, the class of distributions F for which there exists a function f such that

$$F(A) = \int_A f(z)\, d\mu(z) \tag{1.1}$$

We shall call f a generalized probability density function with respect to μ. By Z we denote a random variable such that for any A in $\mathcal{A}$,

$$P[Z \in A] = F(A) \tag{1.2}$$

For most of the applications we shall take $\mathcal{Z}$ to be a Euclidean space, and $\mathcal{A}$ to be the class of all Borel sets. Then if μ is Lebesgue measure, (1.1) states that f is a probability density function in the usual sense. However, we shall have occasion to consider also some measures other than Lebesgue measure. By a hypothesis H we mean a class of probability distributions. Next we describe the hypotheses with which we shall be concerned. Let Π be a partition of $\mathcal{Z}$, that is, let Π be a class of mutually exclusive subsets S of $\mathcal{Z}$ such that every point z of $\mathcal{Z}$ lies in one of the sets S. If two points z_1 and z_2 lie in the same set S, we shall say that z_1 is equivalent to z_2 with respect to Π, $z_1 \sim z_2 \pmod{\Pi}$. The set of all points which are equivalent to z will be denoted by $T(z)$, the number of points of $T(z)$ by $n(z)$. Concerning Π we make the following assumptions.

(i) All sets in Π are finite, so that $n(z)$ is finite for all z.

(ii) If we define S_n as the union of all those sets S of Π which contain exactly n points, there exist mutually exclusive sets $S_n^{(1)}, \cdots, S_n^{(n)}$ which are measurable and such that every element S of Π containing exactly n points has one and only one point in common with each $S_n^{(i)}$.

We shall say that a measure μ is invariant under Π if the following condition holds. For all n and $i, j \leq n$, if S is any set contained in $S_n^{(i)}$ and if S' denotes the set of equivalent points in $S_n^{(j)}$, then $\mu(S) = \mu(S')$.

Given a partition Π satisfying (i) and (ii), we formulate the hypothesis H that the distribution F of Z is invariant under Π. We shall refer to H as the hypothesis of invariance under Π. We shall also consider the hypothesis of invariance under a partition for a class of generalized densities f. In this case we assume that the measure μ of (1.1) is given, and that Π, in addition to (i) and (ii) satisfies the condition

(iii) The measure μ is invariant under Π. The hypothesis H in this case states that $z_1 \sim z_2 \pmod{\Pi}$ implies $f(z_1) = f(z_2)$.

By a test of a hypothesis H we mean (see [4]) a measurable function φ on $\mathcal{Z}$ to the interval $[0, 1]$ which with every point $z \in \mathcal{Z}$ associates a probability $\varphi(z)$ of rejection. This definition, slightly more general than the usual one, is particularly useful in non-parametric work. Among other advantages it automatically takes care of the problem of tied observations. It also disposes of the difficulties encountered by Scheffé [5] in his treatment of the problem of similar regions, as will be shown in Lemma 1.

The size of a test φ is defined to be

$$\varepsilon(\varphi) = \sup_{F \in H} \int \varphi(z)\, dF(z) \tag{1.3}$$

If in particular

$$\int \varphi \, dF = \alpha(\varphi) \tag{1.14}$$

for all F in H, φ is said to be similar for testing H. Extending the terminology of Scheffé, we say that φ has structure $S(\epsilon)$ if for all z in S_n

$$\sum_{z' \in T(z)} \varphi(z') = n\epsilon \tag{1.15}$$

The following lemma extends a result of Scheffé.

LEMMA 1. *For testing a hypothesis of invariance, any test of structure $S(\epsilon)$ is similar and of size ϵ.*

PROOF. For any F in H and any φ

$$\int \varphi \, dF = \sum_{n=1}^{\infty} \sum_{i=1}^{n!} \int_{g_i^{(n)}} \varphi \, dF = \sum_{n=1}^{\infty} \int_{S_n^{(1)}} \Big[\sum_{z' \in T(z)} \varphi(z') \Big] \, dF(z) \tag{1.16}$$

But φ has structure $S(\epsilon)$ and hence (1.5) holds for all z. Therefore

$$\int \varphi \, dF = \sum_{n=1}^{\infty} n\epsilon \int_{S_n^{(1)}} dF = \epsilon. \tag{1.17}$$

We shall show next that for testing a hypothesis of invariance at level of significance ϵ, only tests of structure $S(\epsilon)$ need be considered. In order to make this result applicable both to hypotheses referring to the class of all distributions and to those referring to a class of generalized densities, we shall state it in an asymmetric form which when taken together with lemma 1 indicates the essential equivalence of the two types of hypotheses.

LEMMA 2. *If φ is any test of a hypothesis of invariance for the class of generalized densities with respect to a fixed measure μ, and if the size of φ is less than or equal to ϵ, then there exists a test φ_1 of structure $S(\epsilon)$ such that*

$$\int \varphi_1 \, dF \geq \int \varphi \, dF \tag{1.8}$$

for all probability distributions F.

PROOF. First we shall show that

$$\frac{1}{n(z)} \sum_{z' \in T(z)} \varphi(z') \leq \epsilon \tag{1.9}$$

almost everywhere μ. For let A be the set of points z such that

$$\frac{1}{n(z)} \sum_{z' \in T(z)} \varphi(z') > \epsilon \tag{1.10}$$

and suppose that $\mu(A)$ is positive. Let

$$f(z) = \begin{cases} \dfrac{1}{\mu(A)} & \text{if } z \in A, \\ 0 & \text{elsewhere.} \end{cases} \tag{1.11}$$

Then f is in H since by definition of A, whenever z is in A, $T(z)$ is contained in A. But

$$\int \varphi f \, d\mu > \epsilon \int_A f \, d\mu = \epsilon, \tag{1.12}$$

in contradiction to the assumption that φ has size ϵ.

From (1.9) it follows easily that there exists a test φ_1 of structure $S(\epsilon)$ such that for all z

$$\varphi_1(z) \geq \varphi(z) \tag{1.13}$$

Since condition (1.8) is then satisfied, this completes the proof.

Lemma 2 raises the question whether it is possible to reduce the problem of testing a hypothesis of invariance still further, or whether the tests of structure $S(\epsilon)$ form, what Wald [6] has called an essentially complete class of admissible tests. This question is answered by

THEOREM 1. *Let μ be a measure defined over $\mathcal{Z}$. Let Π_0 and Π_1 be two partitions of $\mathcal{Z}$ satisfying conditions* (i), (ii) *and* (iii), *and such that $z \sim z'$ (mod Π_1) implies $z \sim z'$ (mod Π_0). For the class of generalized densities with respect to μ denote by H_i $(i = 0, 1)$ the hypothesis of invariance relative to Π_i. Then for testing H_0 against H_1 at level of significance ϵ, the totality of tests which (a) have structure $S(\epsilon)$, and for which* (b) $z \sim z'$ (mod Π_1) *implies* $\varphi(z) = \varphi(z')$, *form an essentially complete class of admissible tests.*

PROOF. It is easily seen that we can restrict ourselves to that subclass of tests of structure $S(\epsilon)$ which possess property (b). For if φ is any test of structure $S(\epsilon)$ relative to Π_0, let

$$\varphi^*(z) = \frac{1}{n(z)} \sum_{z' \in T(z)} \varphi(z) \tag{1.14}$$

Then clearly φ^* possesses property (b) and has structure $S(\epsilon)$. Furthermore if f is any probability density function of H_1, then

$$\int \varphi^* f \, d\mu = \int \varphi f \, d\mu, \tag{1.15}$$

so that φ and φ^* have identical power against H_1.

In order to complete the proof, we must show that if φ_1 and φ_2 are any two tests satisfying (a) and (b), and if φ_1 and φ_2 differ on a set of positive measure, there exists a probability density function f of H_1 for which

$$\int \varphi_1 f \, d\mu > \int \varphi_2 f \, d\mu. \tag{1.16}$$

Since both φ_1 and φ_2 have structure $S(\epsilon)$, the set A of points z for which

$$\varphi_1(z) > \varphi_2(z) \tag{1.17}$$

has positive measure. Also, because of (b), if two points are equivalent relative to Π_1, they are either both in A or both not in A. If $f(z)$ is defined as $1/\mu(A)$ for z in A and as zero elsewhere, then f is in H_1 and satisfies (1.16).

The theorem obtained from theorem 1 by letting the hypotheses H_0 and H_1 refer to the class of all probability distributions rather than to a particular class of generalized densities, is clearly also true, and cases between these two theorems could also be formulated.

Since the most powerful test φ for testing a hypothesis of invariance H_0 referring to a class of generalized densities against an alternative f from this class of densities has the correct size also for testing the wider hypothesis H_0 referring to the class of all distributions, φ is also most powerful for testing H_0 against f. The corresponding remark holds for most stringent tests. Therefore all optimum tests that will be derived in the sequel, through the use of theorems of this section, may be considered as tests of hypotheses referring to the class of all distributions that are valid against these hypotheses, and no power is *gained by restricting the hypothesis to the appropriate class of generalized densities.*

2. Most powerful tests and most stringent tests. One of the main problems to be considered in this paper is the determination of a most powerful test of a hypothesis of invariance against a simple alternative. If we restrict our considerations to the class of generalized densities with respect to μ, a complete solution of this problem is given by the following

THEOREM 2. *Let H be the hypothesis of invariance under the partition Π, and let g be a probability density function not in H. For any z in S_n denote by $z^{(1)}, \cdots, z^{(n)}$ the n points of $T(z)$ arranged so that $g(z^{(1)}) \geq g(z^{(2)}) \geq \cdots \geq g(z^{(n)})$. For testing H against g a most powerful test of size ϵ is given by*

$$\varphi(z) = \left\{ \begin{array}{ll} 1 & \text{if } g(z) > g(z^{(1+[n\epsilon])}) \\ a & \text{if } g(z) = g(z^{(1+[n\epsilon])}) \\ 0 & \text{if } g(z) < g(z^{(1+[n\epsilon])}) \end{array} \right\} \quad \text{for } z \text{ in } S_n, \tag{2.1}$$

where $\sum_{i=1}^{n} \varphi(z^{(i)}) = n\epsilon$, $0 \leq a \leq 1$ and where a may depend on z through $T(z)$.

PROOF. First we observe that the number of $z^{(i)}$ for which $g(z^{(i)}) \geq g(z^{(1+[n\epsilon])})$ is greater than or equal to $1+[n\epsilon] \geq n\epsilon$ and that the number of $z^{(i)}$ for which $g(z^{(i)}) > g(z^{(1+[n\epsilon])})$ is less than or equal to $[n\epsilon] \leq n\epsilon$, so that there exists an a between 0 and 1 for which $\Sigma\varphi(z^{(i)}) = n\epsilon$. Since φ has structure $S(\epsilon)$, it follows from lemma 1 that it is similar and of size ϵ.

Let

$$g^*(z) = g(z^{(1+[n\epsilon])}) \quad \text{for} \quad z \in S_n. \tag{2.2}$$

To complete the proof consider first the special case that

$$\int g^*(z)\, d\mu(z) \tag{2.3}$$

vanishes. Then

$$\int \varphi g\, d\mu = \int g\, d\mu = 1 \tag{2.4}$$

that is, the test φ has power 1, and therefore is clearly most powerful. Assume next that the integral (2.3) is positive. Then g^* is proportional to a probability density function of H. For it is measurable and satisfies the measurability condition required of a member of H, and the integral (2.3) is finite since

$$\sum_n \int_{S_n} g^*(z)\, d\mu(z) \le \sum_n \frac{1}{\epsilon n} \int_{S_n} \sum_{i=1}^{[n\epsilon]+1} g(z^{(i)})\, d\mu(z) \tag{2.5}$$

$$\le \frac{1}{\epsilon} \sum_n \int_{S_n} \frac{1}{n} \sum_{i=1}^{n} g(z^{(i)})\, d\mu(z) = \frac{1}{\epsilon}.$$

The test φ therefore has the form of a probability ratio test. Since it is also similar, it follows from theorem 1 of [1] that φ is most powerful.

In practice one is usually interested in composite rather than simple alternatives. We shall therefore consider next the problem of deriving most stringent tests of hypotheses of invariance against certain classes of alternatives. This problem may be reduced to that of finding tests which maximize the minimum power over a class of alternatives by the following simple theorem of Hunt and Stein [7].

THEOREM 3. *Given a hypothesis H and a class of alternatives $\{g_\theta\}$, $\theta \in \Omega$, denote by $\beta^*(\theta)$ the envelope power function corresponding to the level of significance ϵ, that is, let*

$$\beta^*(\theta) = \sup_\varphi \beta(\varphi, \theta) \tag{2.6}$$

where $\beta(\varphi, \theta)$ stands for the power of the test φ against the alternative g_θ and where the least upper bound is taken over all tests φ of size ϵ. Let $\{\Omega_\delta\}$ be a class of mutually exclusive subsets of Ω such that $\bigcup \Omega_\delta = \Omega$ and such that $\beta^*(\theta)$ is constant on each Ω_δ. Denote by φ_δ a test which maximizes the minimum power over Ω_δ. If $\varphi_\delta = \varphi$ is independent of δ, then φ is most stringent[1] for testing H against Ω at level of significance ϵ.

For obtaining tests which maximize the minimum power over a class of alternatives to a hypothesis of invariance, we can state the following simple extension of theorem 2.

THEOREM 4. *Let H be a hypothesis of invariance, and let H_1 be the class of alternatives $\{g_\theta\}$, $\theta \in \Omega$. Suppose there exists a subset Ω' of Ω and a probability measure λ over Ω' such that for the test φ of size ϵ defined as in theorem 2 with*

$$g(z) = \int_{\Omega'} g_\theta(z)\, d\lambda(\theta), \tag{2.7}$$

the integral $\int \varphi g_\theta\, d\mu$ is constant for θ in Ω', and

$$\int \varphi g_\theta\, d\mu \ge \int \varphi g_{\theta'}\, d\mu \quad \text{for all} \quad \theta \in \Omega,\ \theta' \in \Omega' \tag{2.8}$$

Then φ maximizes the minimum power over Ω at level of significance ϵ.

[1] A test is said to be most stringent [10] if it minimizes the maximum difference between envelope power and power, that is, if it minimizes $\sup_\theta [\beta^*(\theta) - \beta(\varphi, \theta)]$.

Proof. By theorem 2, φ is a most powerful test for testing H against g, that is, for any φ' of size ϵ

$$\int \varphi'(z) \int_{\Omega'} g_\theta(z)\, d\lambda(\theta)\, d\mu(z) \leq \int \varphi(z) \int_{\Omega'} g_\theta(z)\, d\lambda(\theta)\, d\mu(z) \tag{2.9}$$

Consequently

$$\begin{aligned} \inf_{\theta \in \Omega} \int \varphi'(z) g_\theta(z)\, d\mu(z) &= \int_{\Omega'} d\lambda(\theta) \int \varphi'(z) g_\theta(z)\, d\mu(z) \\ &= \int \varphi'(z)\, d\mu(z) \int_{\Omega'} g_\theta(z)\, d\lambda(\theta) \leq \int \varphi(z)\, d\mu(z) \int_{\Omega'} g_\theta(z)\, d\lambda(\theta) \\ &= \int_{\Omega'} d\lambda(\theta) \int \varphi(z) g_\theta(z)\, d\mu(z) = \inf_{\theta \in \Omega} \int \varphi(z) g_\theta(z)\, d\mu(z). \end{aligned} \tag{2.10}$$

3. Normal alternatives. Let H be the hypothesis of invariance under Π, let $T(z)$ be the set of points equivalent to z (mod Π), and let f and g be two functions defined over $\mathcal{Z}$. We shall write $f \sim g$ if there exists a function F such that

$$f(z) = F[g(z), T(z)], \tag{3.1}$$

where for any fixed $T(z)$, F is a strictly increasing function of g. We note that $f \sim g$ in the following two special cases:

(i) $f(z) = F[g(z)]$ where F is strictly increasing;

(ii) $f(z) = a(z)g(z) + b(z)$ where $a(z) > 0$ for all z, and where $z_1 \sim z_2$ (mod Π) implies $a(z_1) = a(z_2)$, $b(z_1) = b(z_2)$.

The usefulness of this notation stems from the following remark. Let g^* and φ be defined as in (2.2) and (2.1) respectively and let $f \sim g$. If the test ψ is obtained from φ by substituting f and f^* for g and g^* respectively, then $\psi = \varphi$.

The purpose of the present section is to obtain most powerful and most stringent tests of some hypotheses of invariance against certain classes of normal alternatives. In particular, problems will be exhibited for which various non-parametric tests suggested in the literature possess these optimum properties.

Problem 1. Suppose that the random variables Z_{ij} ($j = 1, \cdots, s_i$; $i = 1, \cdots, m$) have a joint probability density function, and denote by H the hypothesis that this probability density is invariant under all permutations of the s_i arguments within the ith group for $i = 1, \cdots, m$. Consider the alternative H_1 that all variables are independently distributed with common variance σ^2, and that

$$E(Z_{ij}) = ax_{ij} + b_i, \tag{3.2}$$

where a, the b's and the x's are assumed known and where, without essential loss of generality, we assume $a > 0$. Assume further that

$$\sum_{j=1}^{s_i} x_{ij} = 0. \tag{3.3}$$

In order to obtain the most powerful test of H against H_1, we apply theorem 2 with

$$g(z) = c \exp\left[-\frac{1}{2\sigma^2}\Sigma\Sigma(z_{ij} - ax_{ij} - b_i)^2\right] \tag{3.4}$$
$$\sim \Sigma\Sigma(ax_{ij} + b_i)z_{ij} \sim \Sigma\Sigma x_{ij} z_{ij}$$

The most powerful test is therefore given by (2.1), if we replace $q(z)$ by $\Sigma\Sigma x_{ij} z_{ij}$. This test being independent of σ^2, the b's and $a > 0$, it is uniformly most powerful against the class of alternatives obtained from H_1 by not specifying the values of these parameters but restricting a to be positive.

If we drop the restriction $a > 0$, a uniformly most powerful test no longer exists, we shall instead obtain the most stringent test against this extended class of alternatives, using theorems 3 and 4. Clearly the envelope power function is constant on the surfaces $|a|/\sigma^2 = \text{constant}$. Take as the Ω of theorem 1, the set consisting of the two points $(a, b_1, \cdots, b_m, \sigma)$ and $(-a, b_1, \cdots, b_m, \sigma)$. Let λ assign the probability $\frac{1}{2}$ to each of the two points. Then the function g of (2.7) becomes

$$\frac{1}{2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\Sigma s_i} \exp\left\{\frac{1}{2\sigma^2}\Sigma\Sigma(z_{ij} - ax_{ij} - b_i)^2\right\} + \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\Sigma s_i}$$
$$\exp\left\{-\frac{1}{2\sigma^2}\Sigma\Sigma(z_{ij} + ax_{ij} - b_i)^2\right\} \tag{3.5}$$
$$\sim \exp\{\Sigma\Sigma z_{ij}(ax_{ij} + b_i)\} + \exp\{\Sigma\Sigma z_{ij}(-ax_{ij} + b_i)\}$$
$$\sim \exp\{\Sigma\Sigma a x_{ij} z_{ij}\} + \exp\{-\Sigma\Sigma a x_{ij} z_{ij}\} \sim |\Sigma\Sigma x_{ij} z_{ij}|$$

The power of the test φ obtained by substituting this expression for g in (2.1) is the same at both points of Ω. For this test is most powerful for testing H against the simple alternatives H' that the density of the Z's is given by the first member of (3.5). But under the transformation $Z'_{ij} = -Z_{ij} + 2b_i$, H and H' and therefore the test φ are left invariant, while the two points of Ω are permuted.

Condition (2.8) of theorem 4 is therefore satisfied, and hence φ maximizes the minimum power over Ω. Since furthermore φ is independent of the particular set Ω chosen, it follows from theorem 3 that φ is most stringent for the problem under consideration. In case condition (3.3) is not satisfied, let $x'_{ij} = x_{ij} - \bar{x}_i$. Then $\Sigma\Sigma x'_{ij} = 0$ and $E(Z_{ij}) = ax'_{ij} + b'_i$.

Therefore the test criterion (3.5) becomes

$$|\Sigma\Sigma z_{ij}(x_{ij} - \bar{x}_i)| = |\Sigma\Sigma(z_{ij} - \bar{z}_i)(x_{ij} - \bar{x}_i)|. \tag{3.6}$$

Some special cases of problem 1 are of particular interest.

a) Suppose that the variables of the ith group fall into two subgroups, and write for Z_{ij} U_{ij} when $j = 1, \cdots, k_i$, $V_{i,j-k_i}$ when $j = k_i + 1, \cdots, k_i + l_i (k_i + l_i = s_i)$. Let

$$x_{ij} = \begin{cases} 0 & \text{for} \quad j = 1, \cdots, k_i, \\ 1 & \text{for} \quad j = k_i + 1, \cdots, k_i + l_i \end{cases} \tag{3.7}$$

Then the alternatives ascribe to the variables normal distributions with common variance and such that

$$E(U_{ij}) = b_i, \qquad E(V_{ij}) = b_i + a \tag{3.8}$$

The criterion becomes

$$\sum\sum z_{ij}(x_{ij} - x_i) = \sum_{i=1}^{m}\left(\frac{k_i}{k_i + l_i}\sum v_{ij} - \frac{l_i}{k_i + l_i}\sum u_{ij}\right) = \sum \frac{1}{\frac{1}{k_i} + \frac{1}{l_i}}(v_i - u_i) \tag{3.9}$$

or

$$\left|\sum \frac{1}{\frac{1}{k_i} + \frac{1}{l_i}}(v_i - u_i)\right| \tag{3.10}$$

according as a is restricted to positive values or not

b) If we specialize still further and let $m = 1$, we are dealing with a problem which would coincide with the two sample problem if we added independence to the assumptions of the hypothesis (3.10) becomes $|\bar{v} - \bar{u}|$, the criterion suggested by Pitman [3]

c) If instead of m we set $k_i = l_i = 1$ for $i = 1, \cdots, m$ we are testing interchangeability within each pair (u_i, v_i) against normal alternatives under which the means of U_i and V_i are different, the difference being independent of i The criterion $|\sum (v_i - u_i)|$ to which (3.10) reduces was first suggested by R A Fisher [1]

d) As a last example set $m = 1$ in the original problem Under the hypothesis the joint density of $Z_1, \cdots, Z_n$ is symmetric in its n arguments, while under the alternatives the Z's are normally distributed with common variance and mean $ax_i + b$ The criterion reduces to $|\sum (z_i - \bar{z})(x_i - \bar{x})|$ which was proposed by Pitman [3]

We therefore see that several non parametric tests which have been discussed in the literature are most powerful one-sided or most stringent for testing a hypothesis of invariance against certain classes of normal alternatives. In a later section we shall indicate to what extent these results remain valid if to these hypotheses we add the assumption of independence

The remaining problems will be considered somewhat more briefly since the proofs follow the same pattern as in problem 1.

PROBLEM 2 The conditions of problem 1*d*) are satisfied in particular if $x_1, \cdots, x_n$ are values taken on by random variables $X_1, \cdots, X_n$ and if under the alternatives the pairs (X_i, Z_i) have a common bivariate normal distribution with $\sigma_X^2 = \sigma_Z^2$. We are then concerned with a problem related to that of testing for absence of interclass correlation For the corresponding intraclass problem, we consider random variables $X_1, \cdots, X_n, Z_1, \cdots, Z_n$, and test the hypothesis that the joint density of the $2n$ variables is symmetric in all its arguments, against the alternatives that the pairs (X_i, Z_i) have a common bivariate normal distribution, the means and variances of the X's and Z's being the same. We

shall only consider the case of positive correlation. Clearly, the criterion will be $\Sigma x_i z_i$ as in the one sided case of problem (d). However the tests differ, in that this expression must now be compared not only with the $s!$ expressions obtained by permuting the z's among themselves, but instead with the $(2s)!/2^s s!$ expressions obtained by considering all possible ways in which s pairs can be formed from the complete set of $2s$ observations.

PROBLEM 3. Consider once more the hypothesis that the joint density of $Z_1, \cdots, Z_n$ is symmetric in its n arguments, and consider the alternatives that the Z's are normally distributed with positive circular serial correlation. Then

$$(3.11) \qquad g(z) = C \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(z_i - \xi) - \delta(z_{i+1} - \xi)]^2\right\} \sim \sum z_i z_{i+1}$$

where $z_{n+1} = z_1$. The test based on this criterion, which was proposed by Wald and Wolfowitz [8], is therefore most powerful against the above class of alternatives.

PROBLEM 4. As a last problem, we shall test the hypothesis H that the joint density of $Z_1, \cdots, Z_n$ is symmetric in its n arguments and symmetric about each coordinate hyperplane, that is, invariant under the transformation $x_i' = -x_i$, $z_j' = z_j$ for all $j \neq i$, for $i = 1, \cdots, n$. This will be tested against the alternatives that the Z's are independently, identically distributed according to a normal distribution with non-zero mean. If we restrict this mean to positive values, we get

$$(3.12) \qquad g(z) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(z_i - \xi)^2\right\} \sim \Sigma z_i$$

If on the other hand both positive and negative values are allowed for the mean, the most stringent test is based on the statistic $|\Sigma z_i|$.

This test may be appropriate for some situations in which it is customary to use the sign test.

4. Binomial and other non-normal alternatives. In the present section we shall be concerned mainly with generalisations of problems 1b) and 1c) of section 3. As described there, the hypotheses referred to the class of all probability densities in the usual sense. However, as was pointed out at the end of section 2, the same tests may be considered as referring to much wider hypotheses. If they are interpreted in this way, it is possible to greatly widen the class of alternatives without destroying the optimum properties of the tests.

Let $Z = (X_1, \cdots, X_n, Y_1, \cdots, Y_n)$ and denote by Π the partition under which two points z and z' are equivalent if they are obtainable from each other by a permutation of coordinates. Let H_0 be the hypothesis of invariance under Π. This is a generalization of the hypothesis of complete symmetry referring to a class of probability densities. Consider as alternative the class of distributions defined by

$$(4.1) \qquad P\{Z \in A\} = \int_A C \exp\{\theta_1 \Sigma x_i + \theta_2 \Sigma y_i + \Sigma r(x_i) + \Sigma r(y_i)\}\, d\mu(z)$$

where the θ's are any real numbers, where μ is the $2n$th power of any one dimensional measure ν (and therefore invariant under Π), and where r is any ν-measurable function, subject only to the condition that the integral (4.1) converges when taken over the whole space.

We first consider the one sided case $\theta_2 > \theta_1$. Using theorem 2 for a particular θ_1, θ_2, r and μ, we then have

$$g(z) = C' \exp\{\theta_1 \Sigma x_i + \theta_2 \Sigma y_i + \Sigma r(x_i) + \Sigma r(y_i)\}$$

$$\sim \theta_1 \Sigma x_i + \theta_2 \Sigma y_i \sim \theta_1 \Sigma x_i + \theta_2 \Sigma y_i - \tfrac{1}{2}(\theta_1 + \theta_2)\Sigma[x_i + y_i] \tag{4.2}$$

$$= \tfrac{1}{2}(\theta_1 - \theta_2)[\Sigma x_i - \Sigma y_i] \sim \Sigma y_i - \Sigma x_i$$

Since this test does not depend on θ_1, θ_2, r or μ, it is uniformly most powerful against the one sided class of alternatives $\theta_2 > \theta_1$.

Dropping the restriction $\theta_2 > \theta_1$, we apply theorem 4 with Ω the set consisting of the two points θ_1, θ_2, r, μ and θ_2, θ_1, r, μ. At these two points the envelope power function obviously takes on the same value. If for λ we select the distribution, which assigns equal probabilities to both points, then

$$g(z) \sim \exp\{\theta_1 \Sigma x_i + \theta_2 \Sigma y_i\} + \exp\{\theta_2 \Sigma x_i + \theta_1 \Sigma y_i\}$$

$$\sim \exp\{\tfrac{1}{2}(\theta_1 - \theta_2)[\Sigma x_i - \Sigma y_i]\} + \exp\{\tfrac{1}{2}(\theta_2 - \theta_1)[\Sigma x_i - \Sigma y_i]\}$$

$$\sim |\Sigma x_i - \Sigma y_i| \sim |\bar{y} - \bar{x}|$$

The power of this test clearly is the same against both points of Ω. Since furthermore the test does not depend on the θ's, r, or μ, it is most stringent against H_1.

A univariate distribution such that

$$P\{X \in A\} = \int_A C \exp\{\theta x + r(x)\}\, d\nu(x) \tag{4.4}$$

has been called Laplacian by Tweedie [9], who has studied these distributions in a different connection. Among others, the normal and χ^2, the binomial and Poisson distributions are Laplacian. To obtain, for example, the distribution of a characteristic variable, take for ν the measure ν^* which assigns to a set D the values 0, 1 or 2 according as D contains none, one or both of the points $x = 0$ and $x = 1$, and take as density the function

$$p^x(1-p)^{1-x} = (1-p)e^{x \log(p/1-p)} \tag{4.5}$$

For comparison with tests which have been considered in the literature, one can specialize the problem just considered, so that the hypothesis H_0 and the class of alternatives H_1 consist only of those members of H_0 and H_1 which are generalized densities with respect to a fixed measure μ. One can specialize even further and take as alternative any subset of H_1 provided with any point θ_1, θ_2, r, it also contains the point θ_2, θ_1, r. The test clearly will not change with these specializations, and the test based on (4.3) will therefore possess the same

optimum properties with respect to these special problems as with respect to the problem for which it was originally derived.

If in particular one selects for ν the measure ν^* mentioned above, one obtains the problem for which R. A. Fisher proposed the test based on (1.3). It follows that this test, Fisher's exact test, is most stringent in connection with the following problem. The random variables $X_1, \cdots, X_n, Y_1, \cdots, Y_n$ are characteristic variables, that is, they can take on only the values 0 and 1. If we let (4.6) $P\{X_1 = x_1, \cdots, Y_n = y_n\} = P(x_1, \cdots, y_n)$, the hypothesis states that the function P is invariant under all permutations of the arguments. An equivalent formulation is that the probability (4.6) depends only on $\sum x_i + \sum y_i$, the total number of "successes". Fisher's exact test is most stringent against the alternative that the X's and Y's are samples from two distinct populations of characteristic variables, that is, two populations corresponding to distinct probabilities of success.

Problem 1c) of section 3 can be extended quite analogously. Put again $Z = (X_1, \cdots, X_n, Y_1, \cdots, Y_n)$, and denote by Π the partition under which two points z and z' are equivalent provided they can be obtained from each other by a permutation of coordinates in which only the coordinates within pairs (X_i, Y_i) are interchanged. Consider the hypothesis of invariance under Π with reference to the class of all distributions and as alternative the class of distributions given by

$$(4.7) \qquad P\{Z \in A\} = \int_A C \exp\left\{\sum_{i=1}^{n} [\theta_1 x_i + \theta_2 y_i + r(x_i, y_i)]\right\} d\mu(z)$$

The θ's here are any real numbers, μ is the $2n$th power of any one dimensional measure ν, and r is any ν-measurable function such that (a) the integral (4.7) converges when A is the whole space, and such that (b) $r(x, y) = r(y, x)$.

Clearly in the one-sided case $\theta_2 > \theta_1$ we will again find $g(z) \sim \sum y_i - \sum x_i \sim \bar{y} - \bar{x}$, so that the associated test is uniformly most powerful against this one sided class of alternatives, while the test based on $|\bar{y} - \bar{x}|$ is again most stringent against the full alternative H_1.

The class of distributions (4.7) contains the distributions (1.1) as a special case. If (X_i, Y_i) $i = 1, \cdots, n$ is a sample from a bivariate normal distribution with $\sigma_X^2 = \sigma_Y^2$, we get another case of (4.7).

As a last somewhat more special problem we mention a discrete analogue of problem 4 of section 3. Let $Z = (Z_1, \cdots, Z_n)$ and consider the class of generalized densities given by

$$(4.8) \qquad P\{Z \in A\} = \int_A P(z_1, \cdots, z_n)\, d\mu(z)$$

where μ is the nth power of ν^*. Let H_0 be the hypothesis that P is invariant under permutations of the coordinates and under the group generated by the transformations $z_i' = 1 - z_i$, $z_j' = z_j$ $j \neq i$ for $i = 1, \cdots, n$. This is an extension of the hypothesis that the probability of success in a binomial dis-

tribution equals $\frac{1}{2}$. The test of H_0 against the alternatives that $Z_1, \cdots Z_n$ is a sample of a characteristic variable is based on Σz_i or $|\Sigma z_i|$ as $P\{Z_i = 1\}$ is restricted to be greater than $\frac{1}{2}$ or is not so restricted. In the first case the test is most powerful, in the second most stringent.

5. Hypotheses of invariance for independent variables. To the results obtained so far, a different interpretation can be given, which throws some light on certain related problems. Theorem 2 gave sufficient conditions for a test to be most powerful against a simple alternative H_1 for the hypothesis H_0 of invariance under a partition Π. However, if taken in conjunction with section 1, the theorem can be intepreted as giving sufficient conditions for a test to be the most powerful test of structure $S(\epsilon)$ with respect to Π against H_1. That is, the theorem is really independent of the hypothesis, and depends solely on the alternative and on the class of tests admitted into competition, in our case the class of all tests having structure $S(\epsilon)$ with respect to Π. The same remark obviously also applies to most stringent tests.

Let us now consider a special class of partitions. Let Z stand for the m groups of random variables $(Z_{i1}, \cdots, Z_{in})$ $(i = 1, \cdots, m)$ and let Π denote the partition under which two points z and z' are equivalent provided they can be obtained from each other by a permutation of coordinates which however permutes only the coordinates within the m groups. Let μ be the power of a one dimensional measure ν, and assume that the probability distribution of Z is absolutely continuous with respect to μ and that the Z's are independently distributed, so that

$$P\{Z \in A\} = \int_A \prod_{ij} [f_{ij}(z_{ij})\, d\nu(z_{ij})] \tag{5.1}$$

Under these assumptions consider the hypothesis H that f_{ij} is independent of j, that is, that the Z's are identically distributed within each group. It easily can be shown that not all admissible tests of H that have size ϵ, have structure $S(\epsilon)$. However a generalization of a result of Feller [10] and Scheffé [5] for the case $m = 1$ and $\mu =$ Lebesgue measure, states that the only tests which are of size ϵ and similar for H, are the tests of structure $S(\epsilon)$ with respect to Π [11]. It follows that any test which is most powerful or most stringent for testing the hypothesis H' of invariance under Π for the class of generalised densities with respect to μ, has the same property relative to the class of all tests which are similar for testing H.

As an example, take problem 1b) of section 3. Here μ is Lebesgue measure, m is 1, and we put

$$Z_j = \begin{cases} U_j & \text{for } j = 1, \cdots, k \\ V_{j-k} & \text{for } j = k+1, \cdots, k+l = s \end{cases} \tag{5.2}$$

It was shown in section 3 that the test based on $|\bar{u} - \bar{v}|$, Pitman's test, is most stringent for testing the hypothesis that the joint density of the U's and V's is

symmetric in its $k + l$ arguments against the alternative that the variables are independently normally distributed with common variance and such that $E(U_i) = \xi$, $E(V_i) = \eta$ where ξ and η are any distinct real numbers. It follows now that the same test is most stringent similar for testing against the same class of alternatives the hypothesis that $U_1, \cdots, U_k, V_1, \cdots, V_l$ are independently distributed, all with the same probability density. This is the hypothesis for which Pitman proposed his test, and the result just stated is a partial solution of the problem recently raised by Wilks [12], to determine the class of alternatives for which Pitman's test is satisfactory.

If we modify the example by taking for μ instead of Lebesgue measure the $k + l$th power of the measure ν^* of section 4, we are dealing with characteristic variables $U_1, \cdots, U_k, V_1, \cdots, V_l$. We have shown earlier that if $k = l$ the test based on $|\bar{u} - \bar{v}|$ is most stringent for testing the hypothesis of complete permutability against the alternative that the U's and V's are samples from two distinct populations of characteristic variables. If we add to this hypothesis the assumption of independence of all variables, we obtain a parametric problem, namely essentially the problem of testing equality of probability of success in two binomial populations corresponding to the same number of trials. It now follows that the test based on $|\bar{u} - \bar{v}|$ is most stringent for this problem. As is well known, it is also the uniformly most powerful, unbiased similar test.

These two examples suffice to illustrate the type of result that can be obtained. It should perhaps be mentioned that the equivalence discussed at the beginning of this section, can be utilized also in the opposite direction. The fact, for example, that the test based on $|\bar{u} - \bar{v}|$ is known to be uniformly most powerful unbiased similar for testing equality of probability of success in two populations of characteristic variables from which the U's and V's are samples, proves that this test is uniformly most powerful unbiased for testing the hypothesis of complete symmetry for the joint generalized density of the U's and V's.

6. Extension to infinite equivalence classes. The definition of a hypothesis of invariance given in section 1—in spite of the restriction to finite equivalence classes—was sufficiently general to cover the non-parametric problems that we wanted to study. It is possible however to extend the definition so as to allow infinite equivalence classes. In this concluding section we shall briefly outline a theory based on such a broader definition. This will enable us to point out a relationship between the approach of the present paper and the standard parametric theory.

Let $\mathcal{Z}$ be a space of points z and $\mathcal{A}$ an additive class of subsets of $\mathcal{Z}$. We define a partition of $\mathcal{Z}$ into subsets $\{S_t\}$ as follows: Let $\mathcal{T}$ be some space, and for each $t \in \mathcal{T}$ let S_t be a measurable subset of $\mathcal{Z}$ (i.e. an element of $\mathcal{A}$) such that the S_t are mutually exclusive and exhaustive. Let $\mathcal{C}_0$ be the class of all $C_0 \in \mathcal{A}$ which can be expressed in the form

$$C_0 = \bigcup_{t \in D_0} S_t \tag{6.1}$$

and let $\mathcal{D}_0$ be the class of all D_0 occurring in such relationships For each $t \in \mathcal{T}$ let G_t be a specified probability measure over $\mathcal{C}_t$, where $\mathcal{C}_t$ is the class of t_t such that $t_t \in S_t$, $t_t \in \mathcal{C}$ Let Z be a random variable distributed over $\mathcal{Z}$ according to an unknown probability measure F Let $\psi(z)$ be that $t \in \mathcal{T}$ for which $z \in S_t$, and let $T = \psi(Z)$ Let H_0 be the hypothesis that for each $t \in \mathcal{T}$ the conditional distribution of Z given $Z \in S_t$ is G_t, i e that there exists a probability measure Q_0 over $\mathcal{C}_0$ such that for all $A \in \mathcal{C}$

$$F(A) = \int G_t(A \cap S_t)\, dQ_0(t) \tag{6.2}$$

It is seen that we have essentially the situation described in section 1, except that there we assumed further that each S_t was finite and for all t, G_t assigned equal probabilities to all points of S_t.

We say that a test φ of H_0 has structure $S(\epsilon)$ if the conditional expectation $E_t\varphi(Z)$ of $\varphi(Z)$ given $Z \in S_t$ satisfies

$$E_t[\varphi(z)] = \int_{S_t} \varphi\, dG_t = \epsilon \quad \text{for all} \quad t. \tag{6.3}$$

The lemmas and theorems stated below are straight-forward generalizations of those in section 1 so that no proof will be given

LEMMA 1′. *Any test φ of structure $S(\epsilon)$ with respect to H_0 is similar and of size ϵ for H_0.*

LEMMA 2′. *If φ is any test of H_0 of size $\leq \epsilon$, there exists a test φ_1 of H_0 having structure $S(\epsilon)$ and such that*

$$\int \varphi_1\, dF \geq \int \varphi\, dF \tag{6.4}$$

for all probability measures F, for which the conditional distribution of Z given $Z \in S_t$ is absolutely continuous with respect to G_t for all t

Suppose next there is defined another partition of $\mathcal{Z}$ into sets $\{S'_u\}$ by means of a space $\mathcal{U}$, and let $\mathcal{C}'$, $\mathcal{D}'$ and $\mathcal{C}'_u$ refer to this second partition We shall assume that for every $t \in \mathcal{T}$, $u \in \mathcal{U}$ either $S'_u \subset S_t$ or $S'_u \cap S_t$ is empty Let G'_u be a specified probability measure over $\mathcal{C}'_u$ and suppose that for each $t \in \mathcal{T}$ here exists a probability measure Q_t such that for all $A_t \in \mathcal{C}_t$

$$G_t(A_t) = \int G'_u(A_t \cap S'_u) dQ_t(u) \tag{6.5}$$

If H_1 denotes the hypothesis that for each $u \in \mathcal{U}$ the conditional distribution of Z given $Z \in S'_u$ is G'_u, we can state

THEOREM 1′. *For testing H_0 against H_1 at level of significance ϵ, the totality of tests φ which have structure $S(\epsilon)$ and for which $z, z' \in S'_u$ implies $\varphi(z) = \varphi(z')$ form an essentially complete class of admissible tests*

Let F_1 be a distribution not in H_0, and for each $t \in \mathcal{T}$ let G_{1t} be the conditional distribution of Z given $Z \in S_t$ We suppose that for each $t \in \mathcal{T}$, G_{1t} is chosen to be a true probability measure, which is possible in most cases of practical

interest (see Doob [13] for a discussion of this point). Then we have the equivalent of theorem 2:

THEOREM 2'. *Let*

$$G_{1i}(A_i) = \int_{A_i - H_i} g_i \, dG_i + G_{1i}(A_i \cap H_i) \tag{6.6}$$

for all $A_i \subset S_i$, *where in accordance with the Radon-Nikodym Theorem* [14], g_i *is a non-negative function integrable over* S_i, *and* $H_i \subset S_i$ *has* G_i *measure* 0 *and does not depend on* A_i. *For testing* H_0 *against* H_1, *a most powerful test of size* ϵ *is given by* $\varphi(z) = \varphi_i(z)$ *for* $z \in S_i$ *where*

$$\varphi_i(z) = \begin{cases} 1 & \text{if } z \in H_i \\ 1 & \text{if } g_i(z) > c_i \\ a_i & \text{if } g_i(z) = c_i \\ 0 & \text{if } g_i(z) < c_i \end{cases} \tag{6.7}$$

where c_i *and* a_i *are so chosen that* φ *has structure* $S(\epsilon)$.

Theorems 3 and 4 require no modification.

As in the case of finite equivalence classes the results just outlined can be interpreted differently. Again the theorems are really independent of the hypotheses, but depend only on the alternatives and on the class of tests admitted into competition. This class of tests φ is in the present case defined by condition (6.4), that the conditional expectation of φ given $Z \in S_i$ equals ϵ. But this is just the condition which in the standard approach to the problem of testing a composite parametric hypothesis for which T is a sufficient statistic, by means of similar regions is frequently found to be the necessary and sufficient condition for φ to be similar. (See for example [15]). For these cases therefore the hypotheses of the present section represent non-parametric analogues to which the same tests apply with the same optimum properties but without the a priori restriction to similar regions.

As a simple illustration of this remark, let $Z = (Z_1, \cdots, Z_n)$, and let $T = \sum_{i=1}^{n} Z_i^2$. For the conditional distribution of Z given $T = t$ take the uniform distribution over the sphere $T = t$, and for μ take Lebesgue measure. Then the hypothesis H states merely that the joint probability density of the Z's is a function only of $\sum_{i=1}^{n} Z_i^2$. If we add to this the assumption of independence of the Z's, we obtain the new hypothesis H' that the Z's are a sample from a normal distribution with zero mean. The tests φ for which the conditional expectation over each sphere is ϵ, constitute the only admissible tests of H and the only admissible similar tests of H'. If as alternatives we consider that the Z's are a sample from a normal distribution with mean $\xi > 0$, the test

$$\frac{\bar{x}}{\sqrt{\Sigma(x_i - \bar{x})^2}} \gtrless C \tag{6.8}$$

is uniformly most powerful for H and uniformly most powerful similar for H'. If we do not restrict ξ to positive values, the test

$$\left| \frac{\bar{x}}{\sqrt{\Sigma(x_i - \bar{x})^2}} \right| \geq C', \tag{6.9}$$

Student's test, is uniformly most powerful unbiased and most stringent for testing H, uniformly most powerful unbiased similar and most stringent similar for testing H'

REFERENCES

[1] R. A. Fisher, *Design of Experiments*, Oliver and Boyd, Edinburgh, 1935

[2] J. Neyman, K. Iwaszkiewicz and St. Kolodziejczyk, "Statistical problems in agricultural experimentation," *Roy. Stat. Soc. Jour. Suppl.*, Vol. 2 (1935), p. 107

[3] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, *Roy. Stat. Soc. Jour. Suppl.*, Vol. 4 (1937), p. 119; II. The correlation coefficient test, *Roy. Stat. Soc. Jour. Suppl.*, Vol. 4 (1937), p. 225; III. The analysis of variance test, *Biometrika*, Vol. 29 (1938), p. 322

[4] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses. I. Normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948).

[5] H. Scheffé, "On a measure problem arising in the theory of non-parametric tests," *Annals of Math. Stat.*, Vol. 14 (1943), p. 227

[6] A. Wald, "An essentially complete class of admissible decision functions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 549

[7] G. Hunt and C. Stein, "Most stringent tests of statistical hypotheses," *unpublished.*

[8] A. Wald and J. Wolfowitz, "An exact test for randomness in the non-parametric case, based on serial correlation," *Annals of Math. Stat.*, Vol. 14 (1943), p. 378.

[9] M. C. K. Tweedie, "Functions of a statistical variate with given means, with special reference to Laplacian distributions," *Cam. Phil. Soc. Proc.*, Vol. 43 (1947), p. 41

[10] W. Feller, "Note on regions similar to the sample space," *Stat. Res. Memoirs*, Vol. 2 (1938), p. 117

[11] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation," *unpublished*

[12] S. S. Wilks, "Order Statistics," *Am. Math. Soc. Bull.*, Vol. 54 (1948), p. 6.

[13] J. L. Doob, "Asymptotic properties of Markoff transition probabilities," *Trans. Amer. Math. Soc.*, Vol. 63 (1948), footnote p. 399.

[14] S. Saks, *Theory of the Integral*, Stechert, 1937.

[15] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 231 (1933), p. 289

[16] A. Wald, *On the Principles of Statistical Inference*, Notre Dame Mathematical Lectures, Number 1

ESTIMATION OF THE PARAMETERS OF A SINGLE EQUATION IN A COMPLETE SYSTEM OF STOCHASTIC EQUATIONS[1,2]

By T. W. Anderson[3] and Herman Rubin[4]

Columbia University and Institute for Advanced Study

1. Summary. A method is given for estimating the coefficients of a single equation in a complete system of linear stochastic equations (see expression (2.1)), provided that a number of the coefficients of the selected equation are known to be zero. Under the assumption of the knowledge of all variables in the system and the assumption that the disturbances in the equations of the system are normally distributed, point estimates are derived from the regressions of the jointly dependent variables on the predetermined variables (Theorem 1). The vector of the estimates of the coefficients of the jointly dependent variables is the characteristic vector of a matrix involving the regression coefficients and the estimate of the covariance matrix of the residuals from the regression functions. The vector corresponding to the smallest characteristic root is taken. An efficient method of computing these estimates is given in section 7. The asymptotic theory of these estimates is given in a following paper [2].

When the predetermined variables can be considered as fixed, confidence regions for the coefficients can be obtained on the basis of small sample theory (Theorem 3).

A statistical test for the hypothesis of over identification of the single equation can be based on the characteristic root associated with the vector of point estimates (Theorem 2) or on the expression for the small sample confidence region (Theorem 4). This hypothesis is equivalent to the hypothesis that the coefficients assumed to be zero actually are zero. The asymptotic distribution of the criterion is shown in a following paper [2] to be that of χ^2.

2. A complete system of linear difference equations. In many fields of study such as economics, biology, and meteorology the occurrence of values of the observed quantities can be described in terms of a probability model which, as a first approximation, is a set of stochastic equations. Consider a (row) vector y_t of quantities which are observed at time t. Suppose that these quantities are *jointly dependent* on a vector z_t of quantities "predetermined" at time t (i.e., known without error at time t). Some of the coordinates of z_t may be coordinates

[1] This paper will be included in Cowles Commission Papers, New Series, No. 36.

[2] The results in this paper were presented at meetings of the Institute of Mathematical Statistics in Washington, D. C., April 12, 1946 (Washington Chapter) and in Ithaca, N. Y., August 23, 1946.

[3] Fellow of the John Simon Guggenheim Memorial Foundation, Research Consultant of the Cowles Commission for Research in Economics.

[4] National Research Fellow, Research Consultant of the Cowles Commission for Research in Economics.

of y_{t1}, y_{t2}, etc., other coordinates of z_t are quantities which are assumed given constants. The set of vectors $y_t (t = 1, 2, \ldots, T)$ are called *endogenous*. The part of the set z_t which does not consist of lagged endogenous variables is called *exogenous*; these are treated as "fixed variates." For convenience we shall think of t as indicating a point of time, although it may in many cases indicate the ordering of a sample in another dimension, or, indeed, the t may indicate simply a numbering of the observations (if z_t is entirely *exogenous*). In a dynamic economic model the endogenous variables are economic quantities such as amount of investment, interest rate, amount of consumption, etc. The exogenous variables are those quantities which are considered to be determined primarily outside the economic system, such as amount of rainfall, amount of government expenditures, time, etc.

A simple probability model may be set up on the assumption that these quantities approximately satisfy certain linear equations. Specifically the model is

$$B_{vv} y_t' + \Gamma_{vz} z_t' = \epsilon_t' \tag{2.1}$$

where ϵ_t is a (row) vector having a probability distribution with expected value zero and B_{vv} and Γ_{vz} are matrices, the former being non-singular. Primes (′) indicate transposition of vectors and matrices. If there are G jointly dependent variables, there are G component equations in (2.1), that is, there are as many equations as there are variables depending on the system. The fact that y_t and z_t do not satisfy linear equations exactly is indicated by setting the linear forms not equal to zero, but equal to random elements, called *disturbances*. We will call the component equations of (2.1) *structural* equations, for they express the structure of the system. For example, one equation involving the amount of goods consumed, the prices of these goods, the size of the national income, etc., might describe the behaviour of the consumers. Another equation involving interest rate might relate to the behaviour of investors.

It has been shown [7], [11], that in general one cannot use ordinary regression methods to estimate the matrices B_{vv} and Γ_{vz} and the parameters of an assumed distribution of the disturbances. Mann and Wald [9], for a special class of systems, and Koopmans, Rubin, and Leipnik [11], in a more general case, have obtained maximum likelihood estimates of all of the parameters for the case of the ϵ_t having a normal multivariate distribution.

Since B_{vv} is non-singular, we can rewrite (2.1) in a different form, called the *reduced form*,

$$y_t' = -B_{vv}^{-1}\Gamma_{vz} z_t' + B_{vv}^{-1}\epsilon_t', \tag{2.2}$$

or as

$$y_t' = \Pi_{vz} z_t' + \eta_t' \tag{2.3}$$

where

$$\Pi_{vz} = -B_{vv}^{-1}\Gamma_{vz}, \tag{2.4}$$

$$\eta_t' = B_{vv}^{-1}\epsilon_t' \tag{2.5}$$

If ϵ_t has a normal distribution, so does η_t. For a given t then, we can consider the model as specifying a distribution of y_t with conditional expected value $z_t\Pi'_{\cdot i}$.

It is clear that we can multiply (2.1) on the left by any non-singular matrix and obtain a system of equations which defines the same distribution of y_t. On the other hand, it has been shown that the only transformations of $(B_{\nu\nu}\Gamma_{\nu i})$ which preserve the linearity of the system of equations are multiplications on the left by non-singular matrices. If there are a priori restrictions on $(B_{\nu\nu}\Gamma_{\nu i})$, the set of matrices which result in new coefficient matrices satisfying these restrictions is correspondingly decreased. If the set of admissible matrix multipliers includes only diagonal matrices the system of structural equations is said to be *identified*. In this case only multiplication of all coefficients by a given constant is permitted.

Knowledge of the distribution of y_t given z_t is obviously equivalent to knowledge of $\Pi_{\nu i}$ in (2.3) and the distribution of η_t. When the system is identified, the matrix $B_{\nu\nu}$ and

$$\Gamma_{\nu i} = -B_{\nu\nu}\Pi_{\nu i} \tag{2.6}$$

are determined uniquely except for multiplication on the left by a diagonal matrix. Thus identification of a system is equivalent to the possibility of inferring the structural equations from knowledge of the distribution. The estimation of all coefficients of $B_{\nu\nu}$ and $\Gamma_{\nu i}$ has been considered in [11].

3. A single identified equation of a complete system. In many studies the investigator may be interested only in a specific equation of the system, say,

$$\beta_\nu y'_t + \gamma_i z'_t = \zeta_t, \tag{3.1}$$

where ζ_t is a scalar disturbance. The investigator may not be interested in the entire system (2.1) of which (3.1) is one component. Since a considerable amount of computation is necessary to estimate all parameters of a complete system, there arises the problem of estimating only the coefficients of a single equation. It is desirable to do this with the least possible restrictive assumptions about the part of the system which is not the selected structural equation. In order to treat the selected equation at all, we require that it is identified, that is, that there are certain restrictions on (β_ν, γ_i) such that no linear combination of rows of $(B_{\nu\nu}\Gamma_{\nu i})$ satisfies these restrictions other than a constant times (β_ν, γ_i). It is not necessary to assume that every component equation is identified, that is, that the entire system is identified.

We shall suppose that the restrictions imposed are that certain coefficients are zero. We can arrange the components of the vectors so that the restrictions are

$$(\beta_\nu, \gamma_i) = (\beta, 0, \gamma, 0), \tag{3.2}$$

where

$$\beta = (\beta^1, \cdots, \beta^m) \tag{3.3}$$

has H coefficients not assumed to be zero and

$$\gamma = (\gamma^1, \cdots, \gamma^F) \tag{3.4}$$

has F coefficients not assumed to be zero

It will be convenient to divide the G components of y_t into two groups (in number H and $G - H$, respectively), and the K components of z_t into two groups (in number F and D respectively) according to whether or not the components enter into (3.1) with coefficients not assumed to be zero. Let

$$y_t = (x_t, r_t), \tag{3.5}$$

$$z_t = (u_t, v_t), \tag{3.6}$$

where

$$x_t = (x_{t1}, \cdots, x_{tH}), \tag{3.7}$$

$$r_t = (r_{t1}, \cdots, r_{t,G-H}), \tag{3.8}$$

$$u_t = (u_{t1}, \cdots, u_{tF}), \tag{3.9}$$

$$v_t = (v_{t1}, \cdots, v_{tD}). \tag{3.10}$$

Then the selected equation is

$$\beta x_t' + \gamma u_t' = \zeta_t. \tag{3.11}$$

Now let us see how the identification is accomplished. Partitioning Π_{yz} into H and $G - H$ rows and F and D columns as

$$\Pi_{yz} = \begin{pmatrix} \Pi_{xu}^* & \Pi_{xv} \\ \Pi_{ru} & \Pi_{rv} \end{pmatrix},$$

we can write the reduced form (2.3) as

$$x_t' = \Pi_{xu}^* u_t' + \Pi_{xv} v_t' + \delta_t', \tag{3.12}$$

$$r_t' = \Pi_{ru} u_t' + \Pi_{rv} v_t' + \xi_t', \tag{3.13}$$

where

$$\eta_t = (\delta_t, \xi_t).$$

Multiplying the above equation with $(\beta, 0)$ we obtain

$$\beta x_t' = \beta \Pi_{xu}^* u_t' + \beta \Pi_{xv} v_t' + \beta \delta_t'. \tag{3.14}$$

Since this must be identical to (3.11) we must have

$$\gamma = -\beta \Pi_{xu}^*, \tag{3.15}$$

$$0 = -\beta \Pi_{xv}. \tag{3.16}$$

The matrices Π_{xu}^* and Π_{xv} are defined by the distribution of x_t given u_t and v_t (for at least $K = D + F$ linearly independent values of u_t, v_t) The equation

(3 11) is identified if and only if the solution of (3 15) and (3 16) for β and γ is unique except for a constant of proportionality This depends on the rank of Π_{xv} being $H - 1$ Thus a necessary and sufficient condition that (3 11) is identified is that the rank of x_t on v_t be $H - 1$ In particular this implies that the number of coordinates of v_t (the number of zero coefficients in γ_i) be at least $H - 1$ It can easily be shown that this condition is equivalent to requiring that the rank of the matrix obtained by selecting the $G - H$ columns of B_{xy} and the D columns of Γ_{yz} corresponding to the coefficients assumed zero in the selected equation is $G - 1$. This is the condition given by Koopmans and Rubin [11]. Other homogenous linear restrictions can be put in this form.

If the vector ϵ_t is normally distributed with mean zero the vector η_t is normally distributed with mean zero. Let the covariance matrix of δ_t be Ω_{zz} Then the variance of $\zeta_t = \beta\delta_t'$ is

$$\sigma^2 = \beta\Omega_{zz}\beta' \tag{3 17}$$

The constant of proportionality in β may be determined by setting the variance of ζ_t, σ^2, $= 1$, another normalization is

$$\beta^i = 1, \tag{3 18}$$

where β^i is the ith coordinate of β In general the normalization can be written as

$$\beta\Phi_{zz}\beta' = 1, \tag{3.19}$$

where Φ_{zz} can be either a known constant or can be a known function of unknown parameters

As an estimation procedure for β and γ and $D = H - 1$, M A. Girschick suggested in an unpublished note that one solve equations (3 15) and (3 16) with (Π_{zu}^*, Π_{zv}) replaced by $(P_{zu}^* P_{zv})$, the sample regression of x on u and v. By these means Girschick found confidence regions (see section 8) for the parameters of a two equation system A similar idea lies behind a method of O Reiersøl [10]

The present paper develops a method for handling the case of $D \geq H$. In this case the rank of P_{xv} is usually H, thus giving no admissible estimate of β. The proposed method follows the approach used in discriminant problems.

In a second paper [2] the present authors shall give asymptotic properties of these estimates that give a certain justification for the use of them Under very general assumptions concerning the v_t and the ϵ_t we prove that these estimates are consistent. These hypotheses permit the investigator to neglect some predetermined variables absent from his particular equation. Alternative assumptions include the case of the other $G - 1$ equations being non linear Finally, it is shown that the estimates are asymptotically normally distributed For this result it is not necessary to assume that the disturbances are normally distributed, or even that they have identical distributions.

4. A description of the estimation procedure. In a sense the dependence of the endogenous variables x_t on the predetermined variables u_t and v_t is given

by the matrix $(\Pi_{xu}^* \; \Pi_{xv})$ of regression coefficients of x_t on u_t and v_t. The interdependence of the coordinates of x_t indicated by the selected equation nullifies the dependence on z_t, that is,

$$\beta \Pi_{xv} = 0. \tag{4.1}$$

Suppose we wish to estimate β and γ from a sample of T observations: (x_1, z_1), (x_2, z_2), $\cdots$ (x_T, z_T). The information we need can be summarized in the second order moment matrices

$$M_{xx} = \frac{1}{T} \sum_{t=1}^{T} x_t' x_t, \tag{4.2}$$

$$M_{xz} = (M_{xu} M_{xv}) = \frac{1}{T} \left(\sum_{t=1}^{T} x_t' u_t \sum_{t=1}^{T} x_t' v_t \right), \tag{4.3}$$

$$M_{zz} = \begin{pmatrix} M_{uu} & M_{uv} \\ M_{vu} & M_{vv} \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \sum_{t=1}^{T} u_t' u_t & \sum_{t=1}^{T} u_t' v_t \\ \sum_{t=1}^{T} v_t' u_t & \sum_{t=1}^{T} v_t' v_t \end{pmatrix}. \tag{4.4}$$

Since one coordinate of u_t may be unity there is no advantage in taking these moments about the mean. We shall find it more convenient to use instead of v_t the part of v_t that is orthogonal to u_t, that is, we shall use

$$s_t' = v_t' - M_{vu} M_{uu}^{-1} u_t'. \tag{4.5}$$

The moments are then M_{xx}, M_{xu}, M_{uu},

$$M_{xx} = M_{xv} - M_{xu} M_{uu}^{-1} M_{uv}, \tag{4.6}$$

and

$$M_{ss} = M_{vv} - M_{vu} M_{uu}^{-1} M_{uv}. \tag{4.7}$$

We can express the reduced form as

$$x_t' = \Pi_{xu} u_t' + \Pi_{xs} s_t' + \delta_t', \tag{4.8}$$

where

$$\begin{aligned} \Pi_{xu} &= \Pi_{xu}^* + \Pi_{xv} M_{vu} M_{uu}^{-1}, \\ \Pi_{xs} &= \Pi_{xv}. \end{aligned} \tag{4.9}$$

An estimate of Π_{xs} is the regression of x on s,

$$P_{xs} = M_{xs} M_{ss}^{-1}. \tag{4.10}$$

To estimate β we take the β that makes βP_{xs} smallest in the metric determined by the moment matrix of the residuals

$$W_{xx} = M_{xx} - P_{xs} M_{ss} P_{xs}' - P_{xu} \tag{4.11}$$

where

$$P_{xu} = M_{xu}M_{uu}^{-1}. \tag{4.12}$$

This is the natural generalization of least squares, the greatest weight is given to the component with least variance. This estimate is the vector satisfying

$$(P_{xu}M_{uu}P_{xu}' - \nu W_{xx})b' = 0 \tag{4.13}$$

which is associated with the smallest root of

$$|P_{xu}M_{uu}P_{xu}' - \nu W_{xx}| = 0. \tag{4.14}$$

This is normalized and the estimate of γ is $-bP_{xu}$.

In section 5 we derive these estimates by the method of maximum likelihood under certain assumptions. Although it is assumed that the disturbances are normally distributed for this derivation, the estimates can be used in more general situations. This theory is in one sense a special case of the theory of estimating a matrix of means of a given dimensionality which is an extension of the discriminant function theory [5]. For an application of this method of estimation see [6].

5. Derivation of maximum likelihood estimates. We derive the estimates of β, γ, and σ^2 under the following assumptions.

ASSUMPTION A. *The selected structural equation*

$$\beta x_t' + \gamma u_t' = \xi_t \tag{3.11}$$

is one equation of a complete linear system of G stochastic equations. The equation is identified by the fact that if H is the number of coordinates in x_t there are at least $H - 1$ coordinates in v_t, the vector of predetermined variables not in (3.11) but in the system.

ASSUMPTION B. *At time t all of the coordinates of $z_t = (u_t, v_t)$ are given.*

ASSUMPTION C. *The coordinates of z_t are given functions of exogenous variables and of coordinates of $y_{t-1}, y_{t-2}, \cdots$. If coordinates of $y_0, y_{-1}, \cdots$ are involved in z_t, they will be considered as given numbers. The moment matrix M_{zz} is non-singular with probability one.*

ASSUMPTION D. *The disturbance vectors δ_t are distributed serially independently and normally with mean zero and covariance matrix Ω_{zz}.*

We shall consider normalizations (3.10) where Φ_{zz} may be a function of other parameters, but

$$\partial\Phi_{zz}/\partial\beta = 0. \tag{5.1}$$

We can state the results in a theorem.

THEOREM 1. *Under assumptions A, B, C, and D the maximum likelihood estimate of β is*

$$\hat{\beta} = b/\sqrt{b\Phi_{zz}b'}, \tag{5.2}$$

where b is the solution of

$$(P_{zz}M_{zz}P'_{zz} - \nu W_{zz})b' = 0 \tag{4.13}$$

corresponding to the smallest value of ν and P_{zz} is defined by (4.10), M_{zz} *by* (4.6), *and* W_{zz} *by* (4.11). *An estimate of γ based on the maximum likelihood estimate $\hat{\Pi}_{zw}$ is given by*

$$\hat{\gamma} = -\hat{\beta}P_{zw}, \tag{5.3}$$

where P_{zw} is given by (4.12). *The estimate of σ^2 is*

$$\hat{\sigma}^2 = (1 + \nu)/b\hat{\Phi}_{zz}b' \tag{5.4}$$

if

$$bWb' = 1 \tag{5.5}$$

We apply the method of maximum likelihood to

$$L = (2\pi)^{-\frac{1}{2}Tn}\,|\Omega_{zz}^{-1}|^{\frac{1}{2}T}\exp\left\{-\tfrac{1}{2}\sum_{t=1}^{T}(x_t - z_t\Pi'_{zz})\Omega_{zz}^{-1}(x'_t - \Pi_{zz}z'_t)\right. \tag{5.6}$$

under the restrictions (4.1) and (3.19). Replacing v_t by s_t and adding (4.1) and (3.19) multiplied by Lagrange multipliers λ (a vector of D coordinates) and ϕ respectively to the logarithm of L we obtain after division by T

$$\begin{aligned} A = {} & -\tfrac{1}{2}H\log 2\pi + \tfrac{1}{2}\log|\Omega_{zz}^{-1}| + \beta\Pi_{zz}\lambda' + \phi(\beta\Phi_{zz}\beta' - 1) \\ & - \frac{1}{2T}\sum_{t=1}^{T}(x_t - u_t\Pi'_{zu} - s_t\Pi'_{zz})\Omega_{zz}^{-1}(x'_t - \Pi_{zu}u'_t - \Pi_{zz}s'_t) \end{aligned} \tag{5.7}$$

Differentiating (5.7) with respect to β, we obtain

$$\frac{\partial A}{\partial \beta} = \Pi_{zz}\lambda' + 2\phi\Phi_{zz}\beta' \tag{5.8}$$

Setting this equal to zero and multiplying by β, we have

$$\beta\Pi_{zz}\lambda' + 2\phi\beta\Phi_{zz}\beta' = 0.$$

By virtue of (4.1) and (3.19), the Lagrange multiplier ϕ must be zero. Hence, as far as the derivatives of (5.7) are concerned the restriction (3.19) does not enter. The setting of the derivatives of (5.7) equal to zero and (4.1) will define $\hat{\beta}$ except for a constant of proportionality which is finally determined by (3.19). For convenience in deriving the estimates we shall use the normalization

$$\beta\Omega_{zz}\beta' = 1. \tag{5.9}$$

The derivatives of (5.7) with respect to the coordinates of Ω_{zz}, Π_{zw}, Π_{zz}, and β are set equal to zero, resulting in

$$\begin{aligned} \hat{\Omega}_{zz} = {} & M_{zz} - M_{zz}\hat{\Pi}'_{zz} - M_{zw}\hat{\Pi}'_{zw} - \hat{\Pi}_{zz}M_{zz} \\ & - \hat{\Pi}_{zw}M_{wz} + \hat{\Pi}_{zw}M_{ww}\hat{\Pi}'_{zw} + \hat{\Pi}_{zz}M_{zz}\hat{\Pi}'_{zz}, \end{aligned} \tag{5.10}$$

$$\hat{\Omega}_{xx}^{-1}(M_{x1} - \hat{\Pi}_{x1}M_{11}) + \hat{\beta}'\hat{\lambda} = 0, \tag{5.11}$$

$$\hat{\Omega}_{xx}^{-1}(M_{x2} - \hat{\Pi}_{x2}M_{22}) = 0, \tag{5.12}$$

$$\hat{\Pi}_{x1}\hat{\lambda}' = 0 \tag{5.13}$$

Solving (5.12) for $\hat{\Pi}_{x2}$, we obtain

$$\hat{\Pi}_{x2} = P_{x2} \tag{5.14}$$

defined by (4.12). Solving (5.11) for $\hat{\Pi}_{x1}$, we obtain

$$\hat{\Pi}_{x1} = P_{x1} + \hat{\Omega}_{xx}\hat{\beta}'\hat{\lambda}M_{11}^{-1} \tag{5.15}$$

Multiplying (5.15) by $\hat{\beta}$ and solving for $\hat{\lambda}$, we obtain

$$\hat{\lambda} = -\hat{\beta}P_{x1}M_{11}. \tag{5.16}$$

Substitution into (5.15) gives

$$\hat{\Pi}_{x1} = (I - \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta})P_{x1}. \tag{5.17}$$

In view of (5.14) and (5.17) we can write (5.10) as

$$\hat{\Omega}_{xx} = W_{xx} + \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta}P_{x1}M_{11}P_{x1}'\hat{\beta}'\hat{\beta}\hat{\Omega}_{xx}. \tag{5.18}$$

Let

$$\hat{\beta}P_{x1}M_{11}P_{x1}'\hat{\beta}' = \mu. \tag{5.19}$$

Then multiplication of (5.18) on the right by $\hat{\beta}'$ with use of (5.19) gives

$$\begin{aligned}\hat{\Omega}_{xx}\hat{\beta}' &= W_{xx}\hat{\beta}' + \hat{\Omega}_{xx}\hat{\beta}'\hat{\beta}P_{x1}M_{11}P_{x1}'\hat{\beta}' \\ &= W_{xx}\hat{\beta}' + \mu\hat{\Omega}_{xx}\hat{\beta}',\end{aligned}$$

that is,

$$\hat{\Omega}_{xx}\hat{\beta}' = \frac{1}{1-\mu}W_{xx}\hat{\beta}'. \tag{5.20}$$

Equation (5.13) can be written as

$$P_{x1}M_{11}P_{x1}'\hat{\beta}' - \mu\hat{\Omega}_{xx}\hat{\beta}' = 0 \tag{5.21}$$

by substitution from (5.16), (5.17) and (5.19). Combining (5.20) and (5.21) we obtain

$$(P_{x1}M_{11}P_{x1}' - \nu W_{xx})\hat{\beta}' = 0, \tag{5.22}$$

where

$$\nu = \mu/(1-\mu). \tag{5.23}$$

For (5.22) to have a solution, ν must be a root of

$$|P_{x1}M_{11}P_{x1}' - \nu W_{xx}| = 0. \tag{4.14}$$

Substituting from (5.20) into (5.18) we obtain

$$\hat{\Omega}_{xx} = W_{xx} + \mu\left(\frac{1}{1-\mu}\right)^2 W_{xx}\hat{\beta}'\hat{\beta}W_{xx} = W_{xx} + \nu(1+\nu)W_{xx}\hat{\beta}'\hat{\beta}W_{xx}. \tag{5.24}$$

To determine which root of (4.11) to use we shall compute the value of the likelihood function when these estimates are used. It will be convenient to use the solution b of (4.13) with normalization (5.5). Thus b is proportional to $\hat{\beta}$; in fact, since

$$\hat{\beta}\hat{\Omega}_{xx}\hat{\beta}' = \frac{1}{1-\mu}\hat{\beta}W'_{xx}\hat{\beta}'$$

from (5.20), we see that

$$\hat{\beta} = b\sqrt{1-\mu} = b/\sqrt{1+\nu}.$$

Let the other solutions of (4.13) be $b_1, \cdots, b_H$, with corresponding roots $\nu_1, \cdots, \nu_H$, and

$$B^* = \begin{pmatrix} b \\ b_1 \\ \\ \\ b_H \end{pmatrix}$$

Since

$$|\hat{\Omega}_{xx}| = |W_{xx} + \nu W_{xx}b'bW_{xx}|, \tag{5.25}$$

we have

$$|B^*||\hat{\Omega}_{xx}||B^{*\prime}| = |I + \nu B^*W_{xx}b'bW_{xx}B^{*\prime}|. \tag{5.26}$$

Since

$$bW_{xx}B^{*\prime} = (1, 0, \cdots, 0),$$

and since

$$|B^*|^2 = |W_{xx}|^{-1},$$

we deduce from (5.26)

$$|\hat{\Omega}_{xx}| = |W_{xx}|(1+\nu)$$

Multiplying (5.10) by $\hat{\Omega}_{xx}^{-1}$, taking the trace, and substituting in (5.6) we obtain

$$\hat{L} = (2\pi e)^{-\frac{1}{2}TH}|W_{xx}|^{-\frac{1}{2}T}(1+\nu)^{-\frac{1}{2}T}. \tag{5.27}$$

This is a maximum if ν is the smallest root of (4.14).

The theorem now results. The expression for $\hat{\sigma}^2$ follows from

$$\hat{\sigma}^2 = \hat{\beta}\hat{\Omega}_{xx}\hat{\beta}' = b\hat{\Omega}_{xx}b'/b\hat{\Phi}_{xx}b'.$$

If Φ_{xx} is a known constant matrix, $\hat{\Phi}_{xx} = \Phi_{xx}$, if Φ_{xx} is a function of the parameters, $\hat{\Phi}_{xx}$ is the same function of the estimates.

If we define

$$\phi = -\beta\Omega_{xu}^{*}, \tag{5.28}$$

we have by (4.9)

$$\phi = -\hat{\beta}(\Omega_{xu} - \Omega_{xx}M_{xu}M_{uu}^{-1}). \tag{5.29}$$

Since $\hat{\beta}$ annihilates $\hat{\Pi}_{xx}$, (5.3) results
The estimate of Π_{xu} is given by (5.17) and the estimate of Ω_{xx} is

$$\hat{\Omega}_{xx} = W_{xx} + \nu W_{xx}b'bW_{xx}. \tag{5.30}$$

6. The likelihood ratio test of restrictions. It has been assumed that the selected structural equation is identified by imposing the restrictions that certain coefficients are zero. It was noted in Section 3 that at least $G - 1$ such restrictions are necessary. If D, the number of restrictions on the predetermined variables, is more than $H - 1$, we can test the hypothesis that these D coefficients are zero against the alternative that only a smaller number are zero. This is equivalent to a test that Π_{xv} is of rank $H - 1$ against the alternative that the rank is H.

It can be seen intuitively that the smallest root ν of (4.14) indicates how near P_{xx} is to being singular. This statistic can be used to test the hypothesis that Π_{xv} is of rank $H - 1$. The test is similar to the test of rank suggested by P. L. Hsu [8]. The test is stated precisely in the following theorem.

THEOREM 2. *Under assumptions A, B, C, and D the likelihood ratio criterion for testing the hypothesis that Π_{xv} is of rank $H - 1$ against the alternative that it is of rank H is*

$$(1 + \nu)^{-\frac{1}{2}T}, \tag{6.1}$$

where ν is the smallest root of (4.14).

PROOF. If there is no restriction on Π_{xx}, the maximum likelihood estimate of Π_{xx} is P_{xx}, of Π_{xu} is P_{xu}, and of Ω_{xx} is W_{xx}. Then the likelihood function is

$$(2\pi e)^{-\frac{1}{2}TH} \mid W_{xx} \mid^{-\frac{1}{2}T} \tag{6.2}$$

The ratio between this and the likelihood function (5.27) maximized under the hypothesis that the rank of Π_{xv} is $H - 1$ is (6.1).

It is proved in the paper following the present one that under certain conditions (more general than those of Theorem 2)

$$-2 \log [(1 + \nu)^{-\frac{1}{2}T}] = T \log (1 + \nu) \tag{6.3}$$

is distributed asymptotically as χ^2 with $D - H + 1$ degrees of freedom. Thus an approximate test of significance is given by comparing (6.3) with a significance point of the χ^2-distribution with degrees of freedom equal to the excess number of coefficients required to be zero (i.e., the number beyond the minimum required for identification).

7. Computational procedure. The estimation procedure in sections 4 and 5 does not indicate the most efficient method for computing these estimates. The procedure given here is believed to be efficient for ordinary computational equipment and can easily be adapted for sequence-controlled computing machines.

Let us see what expressions occur in the estimation procedure for β and γ. We find that we must first know $P_{xx}M_{qq}P'_{xx}$, W_{xx}, and P_{xy}, these will suffice if Φ_{xx} is constant or Π_{xx} to estimate β, γ, and σ^2. In what follows, we shall assume the normalization is $\beta^1 = 1$, as the results for other normalizations follow immediately. Examining the estimation equations, we see that we may use any matrices proportional to the moment matrices. If equation (3.11) has a constant term, it is better to use moments about the mean and estimate the constant term by setting the calculated mean of the disturbances equal to zero. One possible method of correcting for the mean is to calculate

$$m^*_{pq} = T\sum_{t=1}^{T} p_t q_t - \left(\sum_{t=1}^{T} p_t\right)\left(\sum_{t=1}^{T} q_t\right) \tag{7.1}$$

The estimation procedure for β, σ^2, and the remainder of γ is not affected by correcting for the mean. The computational procedure indicated here is unchanged except for a factor of proportionality in the equation for σ^2 if a different form of correction for the mean is used.

7.1 *Calculation of $M_{xx}M_{zz}^{-1}M_{zx}$ and W_{xx}.* It is known that

$$W_{xx} = M_{xx} - M_{xz}M_{zz}^{-1}M_{zx}. \tag{7.2}$$

We shall use (7.2) to compute W_{xx}. We shall compute $M_{xz}M_{zz}^{-1}M_{zx}$ by the method given by Dwyer [1]. Let us denote the element in the ith row and jth column of M_{zz} by a_{ij}, and the element in the ith row and jth column of M_{zx} by b_{ij}. Let us construct the following array

$$\begin{array}{cccccccc}
c_{11} & c_{12} & \cdots & c_{1K} & c_{11} & c_{12} & \cdots & c_{1H} \\
d_{11} & d_{12} & \cdots & d_{1K} & f_{11} & f_{12} & & f_{1H} \\
 & c_{22} & \cdots & c_{2K} & c_{21} & c_{22} & \cdots & c_{2H} \\
 & d_{22} & & d_{2K} & f_{21} & f_{22} & & f_{2H} \\
 & & \cdots & & & & \cdots & \\
 & & & c_{KK} & c_{K1} & c_{K2} & & c_{KH} \\
 & & & d_{KK} & f_{K1} & f_{K2} & & f_{KH}
\end{array}$$

where

$$c_{ij} = a_{ij} - \sum_{k<i} d_{ki}c_{kj}, \qquad 1 \le i \le j \le h,$$

$$c_{ij} = b_{ij} - \sum_{k<i} d_{ki}c_{kj}, \qquad 1 \le i \le h, 1 \le j \le H$$

$$d_{ij} = \frac{c_{ij}}{c_{ii}}, \qquad 1 \le i \le j \le K,$$

$$f_{ij} = \frac{c_{ij}}{c_{ii}}, \qquad 1 \le i \le K, 1 \le j \le H$$

Then the element in the ith row and jth column of the symmetric matrix $M_{xz}M_{zz}^{-1}M_{zx}$ is

$$\sum_{k=1}^{K} e_{ki} f_{kj} .$$

If we wish to estimate several equations in the system by this method, this step need only be done once, as $M_{xz}M_{zz}^{-1}M_{zx}$ and W_{xx} do not depend upon the equation (except that x would be enlarged)

7.2. *Computation of P_{xu}* We shall compute P_{xu} by the abbreviated Doolittle method Let us now denote the element in the ith row and jth column of M_{uu} by a_{ij}, of M_{uz} by b_{ij} Then let us perform the previous operations, not including the last step We may arrange the work, if only one equation is to be estimated, so that this is already done Then define

$$g_{ij} = f_{ij} - \sum_{i<k\leq F} d_{ik} g_{kj}, \qquad 1 \leq i \leq F, 1 \leq j \leq H.$$

Then the element in the ith row and jth column of P_{xu} is g_{ij}

7 3 *Computation of $P_{xz}M_{zz}P'_{xz}$.* We know that

$$P_{xz}M_{zz}P'_{xz} = M_{xz}M_{zz}^{-1}M_{zx} - M_{xu}M_{uu}^{-1}M_{ux} . \tag{7 3}$$

Let us compute $P_{xz}M_{zz}P'_{xz}$, using (7.3). We must first calculate $M_{xu}M_{uu}^{-1}M_{ux}$. We may do this either by the method of section 7 1, or as $P_{xu}M_{ux}$.

7 4 *Computation of ν, $\hat{\beta}$, and $\hat{\gamma}$.* We shall use

$$\hat{\gamma} = -\hat{\beta} P_{xu} \tag{5 3}$$

to compute $\hat{\gamma}$ after has$\hat{\beta}$ been computed.

Case 1) $H = 1$. In this case the vector $\hat{\beta} = (1)$, $\nu = P_{xz}M_{zz}P'_{xz}/W_{xx}$.

Case 2) $H = 2$, $D > 1$. Let a_{ij} denote the element in the ith row and jth column of $P_{xz}M_{zz}P'_{xz}$, w_{ij} the element in the ith row and jth column of W_{xx}. Define

$$k_0 = |P_{xz}M_{zz}P'_{xz}|,$$

$$k_1 = |W_{xx}|$$

$$k_2 = \tfrac{1}{2}(a_{11}w_{22} + a_{22}w_{11} - 2a_{12}w_{12}).$$

Then

$$\nu = \frac{k_2 - \sqrt{k_2^2 - k_0 k_1}}{k_1},$$

Let $\hat{\Theta} = P_{xz}M_{zz}P'_{xz} - \nu W_{xx}$. Then

$$\hat{\beta}^1 = 1,$$

$$\hat{\beta}^2 = -\frac{\hat{\theta}_{11}}{\hat{\theta}_{12}} = -\frac{\hat{\theta}_{21}}{\hat{\theta}_{22}}$$

Case 3) $H = 2, D = 1$. In this case $\nu = 0$. Then $\hat{\Omega} = P_{xx}M_{xx}P'_{xx}$, and $\hat{\beta}$ may be computed as before.

Case 4) $H > 2, D > H - 1$. Using the procedure of section 7.2, compute $A = (P_{xx}M_{xx}P'_{xx})^{-1}W_{xx}$. Let us multiply equation (5.22) by $-\frac{1}{\nu}(P_{xx}M_{xx}P'_{xx})^{-1}$, and set $1/\nu = \lambda$. We obtain

$$(A - \lambda I)\hat{\beta}' = 0, \tag{7.1}$$

where λ is the largest characteristic root of A. Then we may employ the method of Aitken [1] to estimate λ and $\hat{\beta}$. Let q_0 be an approximation to $\hat{\beta}$. The column of A with largest absolute values is generally a satisfactory approximation. Define

$$q'_i = Aq'_{i-1},$$

$$\lambda_{ij} = \frac{q'_i}{q'_{i-1}}$$

The quantities λ_{ij} approach λ as i increases, and the normalized vectors q_i approach $\hat{\beta}$. The convergence may be accelerated by the methods given by Aitken. The normalization should not be carried out until the λ_{ij} are sufficiently close for different j.

Case 5) $H > 2, D = H - 1$. Let us go through the procedure of section 7.2 with $A = P_{xx}M_{xx}P'_{xx}$, and with no matrix B. Then $c_{HH} = 0$. Set $g_H = 1$, and compute

$$g_i = -\sum_{i<k\leq H} d_{ik}g_k,$$

Then

$$\beta^i = \frac{g_i}{g_1}, \qquad \nu = 0.$$

7.5. *Computation of* $\hat{\sigma}^2$. We have

$$\hat{\sigma}^2 = \hat{\beta}\Omega_{xx}\hat{\beta}' = (1 + \nu)\hat{\beta}W_{xx}\hat{\beta}'. \tag{7.5}$$

If we use the m^*'s instead of the m's, we must divide by T^2, and if other factors of proportionality are used, we must divide by them. σ^2 is in general biased, but the bias depends upon the nature of the complete system, and is not easy to calculate. The bias is of the order of $1/T$.

8. Confidence regions based on small sample theory.[5] If all of the predetermined variables in the system are exogenous (i.e., "fixed"), we can obtain confidence regions for the coefficients of one equation on the basis of small sample theory. To do this we require only that the disturbance of the selected equation be normally distributed; that is, the linear form in the observations $\beta x'_i + \gamma u'_i$

[5] We are indebted to Professor A. Wald for assistance in simplifying our approach to this problem.

is normally distributed with mean zero and variance σ^2 The regression of this on fixed variates is normally distributed and certain quadratic forms in these linear forms have χ^2-distributions On the basis of this we can set up confidence regions for the coefficients

In addition to assumptions A and B we use the following

ASSUMPTION E. *All of the coordinates of* $z_t = (u_t\ v_t)$ *are exogenous The moment matrix* M_{zz} *is non-singular. The disturbances of the selected equation are distributed independently and normally with mean* 0 *and variance* σ^2

Suppose we have a set of observations $(x_1, u_1, v_1), \quad (x_T, u_T, v_T)$ If we know β and γ we can obtain T values of

$$w_t = \beta x_t' + \gamma u_t', \qquad t = 1, \cdots, T. \tag{8.1}$$

The sample regression coefficients of w_t on u_t and s_t are

$$c = \frac{1}{T}\sum_{t=1}^{T} w_t u_t M_{uu}^{-1} = \beta M_{xu} M_{uu}^{-1} + \gamma, \tag{8.1}$$

$$e = \frac{1}{T}\sum_{t=1}^{T} w_t s_t M_{ss}^{-1} = \beta M_{xs} M_{ss}^{-1}. \tag{8.3}$$

The two vectors c and e are distributed independently and normally with mean 0 and covariance matrices

$$\mathcal{E}(c'c) = \sigma^2 M_{uu}^{-1}, \tag{8.4}$$

$$\mathcal{E}(e'e) = \sigma^2 M_{ss}^{-1}, \tag{8.5}$$

Hence (by usual regression theory)

$$C = \frac{1}{\sigma^2} cM_{uu}c' = \frac{1}{\sigma^2}(\beta M_{xu}M_{uu}^{-1}M_{ux}\beta' + \beta M_{xu}\gamma' + \gamma M_{ux}\beta' + \gamma M_{uu}\gamma), \tag{8.6}$$

$$E = \frac{1}{\sigma^2} eM_{ss}e' = \frac{1}{\sigma^2}\beta M_{xs}M_{ss}^{-1}M_{sx}\beta' \tag{8.7}$$

$$= \frac{1}{\sigma^2}\beta(M_{xv} - M_{xu}M_{uu}^{-1}M_{uv})(M_{vv} - M_{vu}M_{uu}^{-1}M_{uv})^{-1}(M_{vx} - M_{vu}M_{uu}^{-1}M_{ux})\beta',$$

$$A = \frac{1}{\sigma^2}\left(\frac{1}{T}\sum_{t=1}^{T} w_t^2 - C - E\right) = \frac{1}{\sigma^2}\beta W_{xx}\beta', \tag{8.8}$$

are distributed independently as χ^2 with F, D, and $T - K$ degrees of freedom, respectively The ratio of any two has an F-distribution

On the basis of these considerations we can obtain the desired confidence regions.

THEOREM 3. *Suppose assumptions A, B, and E are true If the normalization is*

$$\beta\Phi_{xx}\beta' = 1 \tag{8.9}$$

where Ψ_{zz} is a given matrix, (a) a confidence region for β of confidence ϵ consists of all β^* satisfying (8.9) and

$$\frac{\beta^* M_{zz} M_{uu}^{-1} M_{uz} \beta^{*\prime}}{\beta^* W_{zz} \beta^{*\prime}} \frac{T-K}{D} \leq F_{D,T-K}(\epsilon), \tag{8.10}$$

where $F_{D,T-K}(\epsilon)$ is chosen so the probability of (8.10) for $\beta^* = \beta$ is ϵ (b) A confidence region for β and γ simultaneously consists of all β^* and γ^* satisfying (8.9) and

$$\frac{\beta^* M_{zu} M_{uu}^{-1} M_{uz} \beta^{*\prime} + \beta^* M_{zu} \gamma^{*\prime} + \gamma^* M_{uz} \beta^{*\prime} + \gamma^* M_{uu} \gamma^* + \beta^* M_{zz} M_{uu}^{-1} M_{uz} \beta^{*\prime}}{\beta^* W_{zz} \beta^{*\prime}} \tag{8.11}$$

$$\frac{T-K}{K} \leq F_{K,T-K}(\epsilon).$$

(c) If the normalization is $\sigma^2 = 1$, then a confidence region for β of confidence $\epsilon_1 \epsilon_2$ consists of all β^* satisfying

$$\beta^* M_{zz} M_{uu}^{-1} M_{uz} \beta^{*\prime} \leq \chi^2_D(\epsilon_1), \tag{8.12}$$

$$\chi^2_{T-K}(\epsilon_2) \leq \beta^* W_{zz} \beta^{*\prime} \leq \bar{\chi}^2_{T-K}(\epsilon_2), \tag{8.13}$$

where $\chi^2_D(\epsilon_1)$ is chosen so that the probability of (8.12) is ϵ_1 when $\beta^* = \beta$ and $\chi^2_{T-K}(\epsilon_2)$ and $\bar{\chi}^2_{T-K}(\epsilon_2)$ are chosen so that the probability of (8.13) is ϵ_2 when $\beta^* = \beta$ and

$$\underline{\chi}^2(\epsilon_2) \leq 1 \leq \bar{\chi}^2(\epsilon_2) \tag{8.14}$$

(d) A confidence region for β and γ simultaneously consists of all β^* and γ^* satisfying (8.13) and

$$\beta^* M_{zu} M_{uu}^{-1} M_{uz} \beta^{*\prime} + \beta^* M_{zu} \gamma^{*\prime} + \gamma^* M_{uz} \beta^{*\prime} + \gamma^* M_{uu} \gamma^{*\prime} + \beta^* M_{zz} M_{uu}^{-1} M_{uz} \beta^{*\prime} \leq \chi^2_K(\epsilon_1) \tag{8.15}$$

Region (c) is the interior of an ellipsoid and an ellipsoidal shell in the β^*-space, region (d) is similar in the β^*, γ^*-space Region (a) consists of the intersection of the quadric surface (8.9) and the interior of a cone in the β^*-space, region (b) is similar in the β^*, γ^*-space

It is clear that there are many other ways of constructing confidence regions by taking regression on other fixed variates Of these the best seem to be those of theorem 3. It has been proved [2] that the regions of theorem 3 are consistent in the sense that for sufficiently large T the probability is arbitrarily near 1 that all of the confidence region is within a certain distance of β or β, γ For an application of this technique to economic data see a paper by Bartlett [3] who suggested this method independently

9. An approximate small sample test of restrictions. When $\beta^* = \beta$, the probability of (8.10) is ϵ. If β^* is replaced by $\hat{\beta}$ which minimizes the expression

on the left, the probability is at least as great; it is, say, $1 - \delta$. This ratio is λ, the smallest root of

$$\left| \frac{1}{D} M_{xz} M_{zz}^{-1} M_{zx} - \lambda \frac{T}{T - K} W_{xx} \right| = 0, \tag{9.1}$$

Since

$$\lambda = \frac{T - K}{TD} \nu, \tag{9.2}$$

where ν is the smallest root of (4.14), the probability of

$$\nu \geq \frac{TD}{T - K} F_{D, T-K}(\epsilon) \tag{9.3}$$

is $\delta \leq (1 - \epsilon)$. We summarize this as follows.

THEOREM 4 *Under assumptions A, B, and E, the inequality* (9.3), *where* ν *is the smallest root of* (4.14), *constitutes a test of the hypothesis that the coefficients of* y_t *in the selected structural equation are zero of significance less than* $1 - \epsilon$

This test is simply an approximation to the test given in section 6. The exact probability, δ, of (9.3) is unknown; in fact the distribution of ν depends on Π_{z0} and the distribution of δ_t. However, since δ lies between 0 and $1 - \epsilon$, we know that if the test is used as though the level were $1 - \epsilon$, the test will be "conservative."

Another approximate test of the restrictions can be obtained from the inequality (8.11). If the hypothesis is rejected on the basis of one of these tests, the corresponding confidence region (for β or for β and γ) is imaginary, for all β or β and γ are excluded. It should be noticed that the use of a given ratio to test the hypothesis at significance level $\delta (\leq 1 - \epsilon)$ does not affect the confidence coefficient ϵ of the confidence region when the hypothesis is true.

REFERENCES

[1] A. C. AITKEN, "Studies in practical mathematics II. The evaluation of the latent roots and latent vectors of a matrix," *Edinb. Math. Soc. Proc.*, Vol. 57 (1946-7), pp. 269–305.

[2] T. W. ANDERSON AND HERMAN RUBIN, "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations," to be published.

[3] M. S. BARTLETT, "A note on the statistical estimation of demand and supply relations from time series," *Econometrica*, Vol. 16 (1948), pp. 323–329.

[4] P. S. DWYER, "Evaluation of linear forms," *Psychometrika*, Vol. 6 (1941), pp. 355–365.

[5] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol. 8 (1938), pp. 376–386.

[6] M. A. GIRSHICK AND T. HAAVELMO, "Statistical analysis of the demand for food: examples of simultaneous estimation of structural equations," *Econometrica*, Vol. 15 (1947), pp. 79–110.

[7] T. HAAVELMO, "Statistical implications of a system of simultaneous equations," *Econometrica*, Vol. 11 (1943), pp. 1–12.

[8] P. L. Hsu, "On the problem of rank and the limiting distribution of Fisher's test function," *Annals of Eugenics*, Vol. 11 (1941), pp. 39–41.
[9] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp. 173–220.
[10] Olav Reiersøl, "Confluence analysis by means of lag moments and other methods of confluence analysis," *Econometrica*, Vol. 9 (1941), pp. 1–24.
[11] *Statistical Inference in Dynamic Economic Systems*, to be published as Cowles Commission Monograph No. 10.

SOME SIGNIFICANCE TESTS FOR THE MEDIAN WHICH ARE VALID UNDER VERY GENERAL CONDITIONS[1]

By John E. Walsh

The Rand Corporation

1. Summary. Order statistics are used to derive significance tests for the population median which are valid under very general conditions. These tests are approximately as powerful as the Student t test for small samples from a normal population. Also the application of a test requires very little computation. Thus the tests derived compare very favorably with the t-test for small sets of observations. Applications of these order statistic tests to certain well known statistical problems are given in another paper [1].

PART I. RESULTS AND DEFINITIONS

2. Introduction. Consider n independent observations drawn from n populations satisfying the conditions (A):

1) Each population is continuous (i.e. its cdf is continuous).

2) Each population is symmetrical.

3) The median of each population has the same value ϕ. (If the 50% point of a continuous symmetrical population is not unique, the median ϕ of the population is *defined* to be the midpoint of the segment of 50% values.)

It is to be emphasized that no two of the observations are necessarily drawn from the same population. Significance tests are derived to compare ϕ with a given constant value ϕ_0.

A general method of obtaining one sided and symmetrical tests is given in section 8. This general method furnishes tests which have significance levels of the form $r/2^n$, $(r = 1, \cdots, 2^n - 1)$. Each value of r can be attained for some one sided test. Unfortunately tests obtained by the general method are very difficult to apply from a computational viewpoint. If $n \geq 10$, the number of computations required for the application of a test is prohibitive.

To overcome the computational difficulty involved in using the general method, easily applied tests using order statistics are derived. These tests are based on order statistics of certain combinations of order statistics of the n observations, each combination being either a single order statistic of the n observations or one-half the sum of two order statistics. The tests are invariant under permutation of the n observations and have significance levels of the form $r/2^n$, $(r = 1, \cdots, 2^n - 1)$. Table 1 contains a list of some one-sided and symmetrical tests for $n \leq 15$ ($x_1, \cdots, x_n$ represent the n observations arranged in increasing order of magnitude). Additional significance tests can be obtained by use of Theorem 4 of section 6.

---

[1] The results presented in this paper were obtained in the course of research conducted under the sponsorship of the Office of Naval Research. This research was performed while the author was at Princeton University.

If a symmetrical population has a mean, the mean has the same value as the median. Thus if each population from which an observation is drawn satisfies the additional condition that its mean exists, the median tests derived in this paper are also tests of the mean.

Although it is unlikely that conditions (A) are ever exactly satisfied in practice, these conditions appear to be approximately satisfied in many practical situations. Moreover conditions (A) are of such a simple form that approximate verification can frequently be obtained without an extensive investigation.

Certain of the order statistic tests are very efficient if the n observations are a sample from a normal population. Efficiencies are listed for some of the tests in Table I. These tests are approximately as efficient as the Student t-test. (The efficiency of a test, more precisely the power efficiency, is defined in section 3.)

The order statistic tests are competative with the Student t-test. In choosing between the two types of tests the following considerations may be of interest.

(a) The order statistic tests are valid under much more general conditions than the t test.

(b) The order statistic tests are almost as efficient as the t-test for small samples from a normal population.

(c) The order statistic tests are more easily computed than the t-test.

(d) For the case of a sample from a normal population and near significance the t-test gives more information than the order statistic tests.

In some cases a set of n independent observations satisfying only 1) and 3) of conditions (A) can be transformed into observations approximately satisfying all of conditions (A) by an appropriate continuous monotonic change of variable. For example, replacing each observation by the logarithm of the value of the observation sometimes results in a set of observations having approximately symmetrical distributions. Since the transformation, say $g(x)$, is continuous and monotonic, the resulting observations will have median $g(\phi)$ if the original observations have median ϕ. Confidence intervals can be found for ϕ by first obtaining confidence intervals for $g(\phi)$ on the basis of conditions (A) and then inverting. Significance tests can be obtained from these confidence intervals.

The tests of Part I can be applied to furnish generalized solutions for several well known statistical problems. Some of these applications are given in another paper [1].

One application occurs in cases where there is reason to believe that conditions (A) are satisfied but there is no reason to assume that the populations from which the observations were drawn are even approximately the same. Perhaps the most common situation of this type is that in which the value of a certain quantity is experimentally determined by several different methods, all of which should theoretically yield the same result. Then there is no reason to believe that all the experimental values have the same precision. It may be permissible, however, to assume that each value is an observation from a continuous symmetrical population and that all the populations have the same median. Then the order statistic tests can be used to test the true value of the quantity investigated. For example, consider the determination of a specified physical constant

TABLE 1

Some one-sided and symmetrical significance tests for $n \leq 15$

n	Significance Level of Tests		Tests *Symmetrical* Accept $\phi \neq \phi_0$ if either		Approx Efficiency for Normality
	One-sided	Symmetrical	One-sided Accept $\phi < \phi_0$ if	One-sided Accept $\phi > \phi_0$ if	
	%	%			%
4	6 2	12 5	$x_4 < \phi_0$	$x_1 > \phi_0$	95
5	6.2	12 5	$\frac{1}{2}(x_4 + x_5) < \phi_0$	$\frac{1}{2}(x_1 + x_2) > \phi_0$	98
	3 1	6.2	$x_5 < \phi_0$	$x_1 > \phi_0$	96
6	4 7	9 4	$\max[x_5 , \frac{1}{2}(x_4 + x_6)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 + x_3)] > \phi_0$	97
	3 1	6.2	$\frac{1}{2}(x_5 + x_6) < \phi_0$	$\frac{1}{2}(x_1 + x_2) > \phi_0$	98
	1 6	3.1	$x_6 < \phi_0$	$x_1 > \phi_0$	95
7	5 5	10 9	$\max[x_6 , \frac{1}{2}(x_4 + x_7)] < \phi_0$	$\min[x_3 , \frac{1}{2}(x_1 + x_4)] > \phi_0$	95
	2 3	4 7	$\max[x_6 , \frac{1}{2}(x_5 + x_7)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 + x_5)] > \phi_0$	98
	1.6	3 1	$\frac{1}{2}(x_6 + x_7) < \phi_0$	$\frac{1}{2}(x_1 + x_2) > \phi_0$	98
	0 8	1 6	$x_7 < \phi_0$	$x_1 > \phi_0$	95
8	4.3	8 6	$\max[x_6 , \frac{1}{2}(x_4 + x_5)] < \phi_0$	$\min[x_3 , \frac{1}{2}(x_1 + x_3)] > \phi_0$	94 5
	2 7	5 5	$\max[x_6 , \frac{1}{2}(x_5 - x_3)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 + x_4)] > \phi_0$	96
	1.2	2 3	$\max[x_7 , \frac{1}{2}(x_6 - x_5)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 - x_3)] > \phi_0$	98
	0 8	1 6	$\frac{1}{2}(x_7 + x_8) < \phi_0$	$\frac{1}{2}(x_1 + x_2) > \phi_0$	98
	0 4	0.8	$x_8 < \phi_0$	$x_1 > \phi_0$	95
9	5.1	10.2	$\max[x_8 , \frac{1}{2}(x_4 - x_9)] < \phi_0$	$\min[x_4 , \frac{1}{2}(x_1 + x_6)] > \phi_0$	91
	2 2	4 3	$\max[x_7 , \frac{1}{2}(x_5 - x_3)] < \phi_0$	$\min[x_3 , \frac{1}{2}(x_1 - x_5)] > \phi_0$	96
	1 0	2.0	$\max[x_8 , \frac{1}{2}(x_4 - x_9)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 - x_3)] > \phi_0$	97 5
	0.6	1 2	$\max[x_8 , \frac{1}{2}(x_7 - x_6)] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 - x_4)] > \phi_0$	99
	0.4	0 8	$\frac{1}{2}(x_8 + x_9) < \phi_0$	$\frac{1}{2}(x_1 - x_2) > \phi_0$	95

10	5.6	11 1	$\max[x_6 , \frac{1}{2}(x_4 + x_{10})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_7)] > \phi_0$	87 5
	2 5	5 1	$\max[x_7 , \frac{1}{2}(x_5 + x_{10})] < \phi_0$	$\min[x_4 , \frac{1}{2}(x_1 + x_6)] > \phi_0$	93
	1.1	2.1	$\max[x_8 , \frac{1}{2}(x_6 + x_{10})] < \phi_0$	$\min[x_3 , \frac{1}{2}(x_1 + x_5)] > \phi_0$	96 5
	0.5	1 0	$\max[x_9 , \frac{1}{2}(x_6 + x_{10})] < \phi_0$	$\min[x_2 , \frac{1}{2}(x_1 + x_5)] > \phi_0$	96 5
11	4.8	9.7	$\max[x_7 , \frac{1}{2}(x_4 + x_{11})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_8)] > \phi_0$	
	2.8	5.6	$\max[x_7 , \frac{1}{2}(x_5 + x_{11})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_7)] > \phi_0$	89
	1 1	2 1	$\max[\frac{1}{2}(x_6 + x_{11}), \frac{1}{2}(x_8 + x_9)] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_6), \frac{1}{2}(x_3 + x_4)] > \phi_0$	
	0.5	1.1	$\max[x_9 , \frac{1}{2}(x_7 + x_{11})] < \phi_0$	$\min[x_3 , \frac{1}{2}(x_1 + x_5)] > \phi_0$	97
12	4.7	9.4	$\max[\frac{1}{2}(x_4 + x_{12}), \frac{1}{2}(x_5 + x_{11})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_9), \frac{1}{2}(x_2 + x_8)] > \phi_0$	
	2.4	4.8	$\max[x_8 , \frac{1}{2}(x_5 + x_{12})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_9)] > \phi_0$	
	1 0	2.0	$\max[x_9 , \frac{1}{2}(x_6 + x_{12})] < \phi_0$	$\min[x_4 , \frac{1}{2}(x_1 + x_7)] > \phi_0$	93 5
	0 5	1.1	$\max[\frac{1}{2}(x_7 + x_{12}), \frac{1}{2}(x_9 + x_{10})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_6), \frac{1}{2}(x_3 + x_4)] > \phi_0$	
13	4 7	9.4	$\max[\frac{1}{2}(x_4 + x_{13}), \frac{1}{2}(x_5 + x_{12})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{10}), \frac{1}{2}(x_2 + x_9)] > \phi_0$	
	2.3	4 7	$\max[\frac{1}{2}(x_5 + x_{13}), \frac{1}{2}(x_6 + x_{12})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_9), \frac{1}{2}(x_2 + x_8)] > \phi_0$	
	1 0	2.0	$\max[\frac{1}{2}(x_6 + x_{12}), \frac{1}{2}(x_9 + x_{10})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_8), \frac{1}{2}(x_4 + x_5)] > \phi_0$	
	0.5	1 0	$\max[x_{10} , \frac{1}{2}(x_7 + x_{13})] < \phi_0$	$\min[x_4 , \frac{1}{2}(x_1 + x_7)] > \phi_0$	94 5
14	4 7	9.4	$\max[\frac{1}{2}(x_4 + x_{14}), \frac{1}{2}(x_5 + x_{13})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{11}), \frac{1}{2}(x_2 + x_{10})] > \phi_0$	
	2 3	4 7	$\max[\frac{1}{2}(x_6 + x_{14}), \frac{1}{2}(x_6 + x_{13})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{10}), \frac{1}{2}(x_2 + x_9)] > \phi_0$	
	1 0	2.0	$\max[x_{10} , \frac{1}{2}(x_8 + x_{14})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_9)] > \phi_0$	90 5
	0 5	1 0	$\max[\frac{1}{2}(x_7 + x_{14}), \frac{1}{2}(x_{10} + x_{11})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_8), \frac{1}{2}(x_4 + x_6)] > \phi_0$	
15	4 7	9 4	$\max[\frac{1}{2}(x_4 + x_{15}), \frac{1}{2}(x_5 + x_{14})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{12}), \frac{1}{2}(x_2 + x_{11})] > \phi_0$	
	2 3	4.7	$\max[\frac{1}{2}(x_6 + x_{15}), \frac{1}{2}(x_6 + x_{14})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{11}), \frac{1}{2}(x_2 + x_{10})] > \phi_0$	
	1 0	2 0	$\max[\frac{1}{2}(x_6 + x_{15}), \frac{1}{2}(x_{10} + x_{11})] < \phi_0$	$\min[\frac{1}{2}(x_1 + x_{10}), \frac{1}{2}(x_5 + x_6)] > \phi_0$	
	0 5	1 0	$\max[x_{11}, \frac{1}{2}(x_7 + x_{15})] < \phi_0$	$\min[x_5 , \frac{1}{2}(x_1 + x_9)] > \phi_0$	92

Various scientists obtained experimental values for this constant by several different methods If it can be assumed that each value is an observation from a continuous symmetrical *population and* that all the populations have the same median, the true value of the physical constant can be tested by applying the order statistic tests to the totality of the experimental values

3 Power efficiency of tests. A problem which arises throughout the paper is that of determining how much *information* is lost by using some other test in place of the most powerful test of a given hypothesis The quantitative measure of the amount of available information which is used by a test will be given as a percentage and is called the power efficiency of the test considered

In all cases investigated the underlying population is normal with unknown variance and the hypotheses tested concern the *population median (mean)* Then the most powerful test (one-sided or symmetrical) is the appropriate Student t-test

The procedure used to measure the power efficiency of a test is different from the common method of measuring the efficiency of an estimate The efficiency of an estimate is obtained by *taking the ratio of the variance of an efficient estimate* with respect to the variance of the given estimate (expressed as a percentage) The method of determining the power efficiency of a test, however, consists in continuously varying the sample size of the appropriate most powerful test (same significance level) until the power functions of the given test and the most powerful test are equivalent in the following sense: The area between the two power curves for which the power function of the most powerful test exceeds the power function of the given test is equal to the analogous area for which the power function of the most powerful test is less than that of the given test (It is assumed that the power functions of the tests can be made to depend on the values of a single parameter) *The sample size (not necessarily integral) of the most powerful test with equivalent power function divided by the sample size of the given test is called the power efficiency of the given test (expressed as a percentage)*

In obtaining power efficiencies in the manner defined above, the sample size of the most powerful test is allowed to assume non-integral values This furnishes an interpolated measure of the same size of the most powerful test which is power function equivalent to the given test As pointed out above, the t test is a most powerful test for the situations considered in this paper A method of computing power function values for t-tests having non-integral sample sizes is given below

The definition of power efficiency selected is very convenient from a computational point of view Power function values for the t test can be easily computed through use of the normal approximation given in [2]. For the significance levels considered in this paper, the normal approximation is reasonably accurate if the sample size is not too small In the remaining cases the approximation underestimates some power function values and overestimates others For the situations investigated, however, the error introduced by this combination of

TABLE 2

Efficiences and power function values for certain order statistic tests

Significance Test	Sample Size	Approx Efficiency	Significance Level	Values of Power Function		
				$\delta = 0$	$\delta = 1.2$	$\delta = 1.8$
		%				
l	4 9		.0625	.337	.755	.964
$\frac{1}{2}(x_4 + x_6) < \phi_0$	5	98	.0625	.343	.755	.958
l	5 82		.0160	.327	.779	.980
$\max[x_5, \frac{1}{2}(x_4 + x_6)] < \phi_0$	6	97	.0469	.334	.779	.972
l	5 88		.0312	.244	.682	.951
$\frac{1}{2}(x_5 + x_6) < \phi_0$	6	98	.0312	.254	.687	.942
l	6 65		.0547	.406	.809	.994
$\max[x_5, \frac{1}{2}(x_4 + x_7)] < \phi_0$	7	95	.0547	.413	.867	.991
l	6 85		.0231	.239	.716	.969
$\max[x_6, \frac{1}{2}(x_5 + x_7)] < \phi_0$	7	98	.0231	.249	.717	.962
l	7 55		.0130	.395	.882	.996
$\max[x_6, \frac{1}{2}(x_4 + x_8)] < \phi_0$	8	94 5	.0130	.401	.879	.993
l	7 85		.0117	.174	.650	.956
$\max[x_7, \frac{1}{2}(x_6 + x_8)] < \phi_0$	8	98	.0117	.185	.656	.949
l	8.61		.0215	.302	.839	.994
$\max[x_7, \frac{1}{2}(x_5 + x_9)] < \phi_0$	9	96	.0215	.311	.834	.990
l	8 9		.0059	.127	.597	.947
$\max[x_8, \frac{1}{2}(x_7 + x_9)] < \phi_0$	9	99	.0059	.137	.599	.935
l	7 5		.0547	.450	.910	.998
$x_8 < \phi_0$	10	75	.0547	.454	.901	.995
l	9 65		.0107	.227	.790	.991
$\max[x_8, \frac{1}{2}(x_6 + x_{10})] < \phi_0$	10	90.5	.0107	.237	.786	.986
l	8 2		.0098	.176	.068	.964
$\max[x_9, \frac{1}{2}(x_1 + x_{10})] < \phi_0$	10	82	.0098	.191	.077	.952
l	8.9		.0059	.141	.621	.954
$x_{10} < \phi_0$	11	81	.0059	.152	.631	.942
l	11.22		.0102	.277	.870	.998
$\max[x_9, \frac{1}{2}(x_6 + x_{12})] < \phi_0$	12	93.5	.0102	.288	.862	.995

underestimation and overestimation tends to cancel out in the determination of power efficiencies if the above area definition of equality of power functions is used Thus application of the normal approximation yields reasonably accurate power efficiencies for the cases considered in this paper Use of the normal approximation furnishes an easily applied method of obtaining power function values for t-tests having non-integral sample sizes

Table 2 contains examples of the above described method of determining power efficiencies Here the power function values for the t test were computed using the normal approximation Examination of Table 2 shows that the maximum difference between corresponding power function values for the two types of tests is small for all the cases considered there This holds in the determination of all the power efficiencies listed in Table 1

Investigation indicates that the definition of power efficiency given here is for all practical purposes the same as that given in [3]

For the situations considered in this paper, it is sufficient to restrict power efficiency investigations to one-sided tests. Every symmetric test investigated can be considered as a combination of two non overlapping one-sided tests, each having a significance level equal to half that of the symmetric test Also, from symmetry, these one-sided tests (each considered as a separate test) have the same power efficiency. Thus it is an immediate consequence of the definition of power efficiency that the symmetric test has the same efficiency as each of the corresponding one-sided tests at half the significance level

PART II DERIVATIONS

4. Introduction. The purpose of the remainder of the paper is to present derivations of the significance test results stated in sections 1 and 2 The first derivations consist in obtaining confidence intervals for ϕ on the basis of conditions (A). Then properties of these confidence intervals are analyzed Application of the confidence intervals and then properties to significance tests furnishes many of the results stated in sections 1 and 2 The remaining derivations are concerned with efficiencies and the general method mentioned in section 2.

5. Derivation of confidence intervals. Let us consider n independent observations, each observation being drawn from a possibly different population Denote these observations by $y_1, \quad , y_n$ and let the cdf of y_i be given by F_i, $(i = 1, \quad , n)$ Furthermore let the n populations from which these n observations were drawn satisfy conditions (A). Then 1) of conditions (A) requires that each F_i is continuous, while 2) and 3) stipulate that

$$\int_{-\infty}^{-c} dF_i(y_i - \phi) = \int_{c}^{\infty} dF_i(y_i - \phi), \qquad (i = 1, \quad , n),$$

for all values of c in the interval $-\infty < c < \infty$

Let $x_1, \quad , x_n$ represent $y_1, \quad , y_n$ arranged in increasing order of magnitude. Since the cdf's are continuous, $Pr(x_i = x_j; i \neq j) = 0$ For the situa-

tions treated in this paper, it is sufficient to consider one-sided confidence intervals for ϕ. All one sided confidence intervals derived have one of the forms

$$(1) \qquad \begin{aligned} g(x_1, \cdots, x_n) &< \phi, \\ h(x_1, \cdots, x_n) &> \phi, \end{aligned}$$

where g and h are Borel measurable functions of $x_1, \cdots, x_n$ such that

$$Pr[g(x_1, \cdots, x_n) < \phi] = Pr[g(x_1 - \phi, \cdots, x_n - \phi) < 0],$$
$$Pr[h(x_1, \cdots, x_n) > \phi] = Pr[h(x_1 - \phi, \cdots, x_n - \phi) > 0].$$

Consider the additional condition

(B) All populations are the same

In terms of cumulative distribution functions, condition (B) requires that all the cdf's F_i are equal to some cdf F. A theorem will be proved which shows that all confidence intervals of the forms (1) derived on the basis of both conditions (A) and (B) are also valid if only conditions (A) necessarily hold, i.e. if

$$Pr[g(x_1, \cdots, x_n) < \phi] = p$$

whenever $x_1, \cdots, x_n$ are order statistics of observations from populations satisfying conditions (A) and (B), then this probability expression also has the value p if $x_1, \cdots, x_n$ are from populations necessarily satisfying only conditions (A). Similarly for $Pr[h(x_1, \cdots, x_n) > \phi]$.

THEOREM 1. *Let $Q(x_1 - \phi, \cdots, x_n - \phi)$ be a probability statement involving $x_1 - \phi, \cdots, x_n - \phi$ which defines a Borel measurable region $R(x_1 - \phi, \cdots, x_n - \phi)$ of the n-dimensional order statistic space. If*

$$(2) \qquad Q(x_1 - \phi, \cdots, x_n - \phi) = p$$

whenever $x_1, \cdots, x_n$ are order statistics of n independent observations from populations satisfying conditions (A) and (B), then (2) also holds when $x_1, \cdots, x_n$ are order statistics of n independent observations from populations necessarily satisfying only conditions (A).

PROOF. It is sufficient to consider the case in which $\phi = 0$. Then, if conditions (A) are satisfied, the joint probability element of $x_1, \cdots, x_n$ is

$$dF(x_1, \cdots, x_n) = \sum_{\pi} dF_1(x_{\pi(1)}) \cdots dF_n(x_{\pi(n)}),$$

where the summation is taken over all permutations π of the integers $1, \cdots, n$, and F's are cdf's of symmetrical populations with zero median. Let $R = R(x_1, \cdots, x_n)$ be the region of the n-dimensional order statistic space defined by the probability statement $Q(x_1, \cdots, x_n)$. Then Theorem 1 stipulates that

$$(3) \qquad \int_R dF(x_1, \cdots, x_n) = p$$

whenever $y_1, \cdots, y_n$ are from populations satisfying conditions (A) and (B) with zero median. In this case, however, each $F_i = F$ and (3) becomes

$$n! \int_R \prod_{i=1}^{n} dF(x_i) = p, \tag{4}$$

where F is the cdf of a population satisfying conditions (A) and (B) with zero median. Let

$$P = \prod_{i=1}^{n} \left(\sum_{j=1}^{n} dF_j(x_i) \right)$$

and define S_α^β to be the sum of all terms in the expansion of P which contain a specified α of $dF_1, \cdots, dF_n$, and no others, the particular set chosen is denoted by β, where $\beta = 1, \cdots, \binom{n}{\alpha}$. Then

$$P = F(x_1, \cdots, x_n) + \sum_{\beta} S_{n-1}^{\beta} + \cdots + \sum_{\beta} S_1^{\beta}$$

Now consider any given S_α^β (i.e. α, β given). Define dH to be the sum of the α of $dF_1, \cdots, dF_n$ pertaining to β plus any set of zero or more of the remaining dF's. Then no matter which of the remaining dF's are chosen for dH, the sum of those terms in the expansion of $\prod_{i=1}^{n} dH(x_i)$ which contain the particular set of α of $dF_1, \cdots, dF_n$ is always equal to S_α^β. Let

$$P_\alpha = \sum_{\beta} \left(\prod_{i=1}^{n} dG_\alpha^\beta(x_i) \right),$$

where dG_α^β equals the sum of the α of $dF_1, \cdots, dF_n$ pertaining to β. Then from the above and the symmetrical fashion in which the dF's are treated,

$$P_\alpha = \sum_{\beta} S_\alpha^\beta + K_{\alpha-1}^{(\alpha)} \sum_{\beta} S_{\alpha-1}^{\beta} + \cdots + K_1^{(\alpha)} \sum_{\beta} S_1^{\beta},$$

where the $K_u^{(\alpha)}$ $(u = 1, \cdots, \alpha - 1)$, are constants.

Consider the case in which $\alpha = n - 1$. Using the above expression for P_{n-1},

$$P = dF(x_1, \cdots, x_n) + P_{n-1} + (1 - K_{(n-2)}^{(n-1)}) \sum_{\beta} S_{n-2}^{\beta} + \cdots + (1 - K_1^{(n-1)}) \sum S_1^{\beta}$$

Repeating this procedure successively for $\alpha = n - 2, n - 3, \cdots, 1$ shows that

$$dF(x_1, \cdots, x_n) = P + C_{n-1} P_{n-1} + \cdots + C_1 P_1,$$

where the C_v, $(v = 1, \cdots, n - 1)$, are constants.

Since each F_i is the cdf of a symmetrical population with zero median,

$$G_\alpha^\beta/\alpha = \tfrac{1}{\alpha} \text{ (sum of the } \alpha \text{ of } F_1, \cdots, F_n \text{ pertaining to } \beta)$$

is also the cdf for a continuous symmetrical population with zero median But

$$F'_\alpha - \alpha^n \binom{F'_\alpha}{\alpha^n} = \alpha^n \sum_\beta \left(\prod_{i=1}^{n} d(\tau_\alpha^\beta(\iota_i)/\alpha \right)$$

Hence $dF(x_1, \quad , x_n)$ is equal to a sum of terms (multiplied by certain constants) of the form

$$n! \prod_{i=1}^{n} dF(\varepsilon_i),$$

where F is the cdf of a continuous symmetrical population with zero median. Thus from (1) and the linear properties of the integral,

$$\int_R dF(x_1, \cdot \quad , x_n) = p$$

if $\eta_1, \quad , \eta_n$ are from populations necessarily satisfying only conditions (A) Q e d

Next confidence intervals of the forms (1) will be derived for ϕ on the basis of conditions (A) and (B). Before stating the theorem on which these confidence intervals are based consider the following definition of notation: For each permissible selection of i and j the symbol

$$[r, j] \qquad (1 \leq i \leq j \leq n)$$

denotes an *arbitrary* but *fixed* selection of one or both of the inequality signs $<, >$ The selection of both inequality signs, denoted by $\lesssim$, has the interpretation

$$r_i \lesssim \phi = -\infty < r_i < \infty$$

$$(x_i + x_j)/2 \lesssim \phi = -\infty < (v_i + v_j)/2 < \infty$$

It is to be noted that $[r, s]$ is not necessarily equal to $[i, j]$ unless $r = i$ and $s = j$

THEOREM 2 *Consider the probability statement*

$$Pr\{(v_i + v_j)/2 \; [i, j] \; \phi; \; 1 \leq i \leq j \leq n\} \tag{5}$$

Let this statement have the value q if $v_1, \cdot \quad , v_n$ are order statistics of a sample of size n drawn from the uniform population with range $-\frac{1}{2}$ to $\frac{1}{2}$ (then $\phi = 0$) Then (5) also has the value q if $x_1, \quad , v_n$ are order statistics of a sample size n drawn from any population satisfying conditions (A) and (B)

PROOF Let $y_1, \cdot \quad , y_n$ be a sample of n values from a population satisfying conditions (A) and (B) while $v_1, \quad \cdot, v_n$ are the y's arranged in increasing order of magnitude. Then there is a monotone function π (see [1]) such that $\pi(z)$ will have the same cdf as $y_i - \phi$ if z is from a uniform population with range $-\frac{1}{2}$ to $\frac{1}{2}$ Since the y's are from a symmetrical population, $-\pi(z) = \pi(-z)$. Let $v_i - \phi = \pi(z_i)$, $(i = 1, \cdot \cdot, n)$, define the z_i Then

$$Pr[(x_i + x_j)/2 \{i, j\} \phi] = Pr[(\pi(z_i) + \pi(z_j)) \{i, j\} 0]$$
$$= Pr[\pi(z_i) \{i, j\} - \pi(z_j)]$$

From the monotone and symmetrical properties of the function π,

$$Pr[\pi(z_i) \{i, j\} - \pi(z_j)] = Pr[\pi(z_i) \{i, j\} \pi(-z_i)]$$
$$= Pr[z_i \{i, j\} - z_i]$$

By hypothesis this last expression has the value q, thus completing the proof

Many of the probability statements of the form (5) have zero probability For example, $Pr[x_1 > \phi, x_2 < \phi, \cdot \quad] = 0$ Also many selections of the symbols $\{i, j\}$ result in equivalent probability statements For example

$$Pr(x_1 \lesseqgtr \phi, x_2 < \phi) \equiv Pr(x_1 < \phi, x_2 < \phi).$$

An immediate consequence of Theorem 2 is that one-sided confidence intervals can be obtained for ϕ by choosing any specified subset of $(x_i + x_j)/2$, $(1 \leq i \leq j \leq n)$, and considering an arbitrary but fixed order statistic of the values of this subset For example, consider the subset consisting of x_{n-1} and $(x_{n-2} + x_n)/2$. Then

$$Pr[\max[x_{n-1}, (x_{n-2} + x_n)/2] < \phi] = Pr[(x_i + x_j)/2 \{i, j\} \phi],$$

where

$$\{i, j\} = \begin{cases} < \text{ if either } i = j = n - 1; \quad \text{or } i = n - 2, \quad j = n, \\ \lesseqgtr \text{ otherwise.} \end{cases}$$

In general, the confidence coefficient of any one-sided confidence interval formed by considering a certain order statistic of a specified subset of $(x_i + x_j)/2$, $(1 \leq i \leq j \leq n)$, can be expressed as a sum of probabilities of the form (5), where $\{i, j\} = \lesseqgtr$ if $(x_i + x_j)/2$ is not included in the specified subset, $(i < j)$

It is usually preferable to select the subset of $(x_i + x_j)/2$, $(1 \leq i \leq j \leq n)$, in such a way that no two of the elements chosen necessarily have an order relation

Satisfactory two sided confidence intervals can usually be obtained as combinations of one-sided confidence intervals

6 Confidence coefficients. The purpose of this section is to show that all the confidence coefficients for one-sided confidence intervals derived on the basis of Theorem 2 are of the form $r/2^n$, $(r = 1, \quad , 2^n - 1)$. Also a method of determining confidence coefficient values for one-sided confidence intervals is developed.

First a theorem will be presented which shows that each of the one sided confidence intervals derived in the preceding section has a confidence coefficient of the form $r/2^n$, $(r = 1, \cdots, 2^n - 1)$. On the basis of Theorem 2 it is sufficient to prove,

THEOREM 3 *Let* $x_1, \quad , x_n$ *be the ordered values of a sample from the uniform population with range* $-\frac{1}{2}$ *to* $\frac{1}{2}$ *Then*

$$Pr[(x_i + x_j)/2\{i, j\}\ 0, 1 \leq i \leq j \leq n] = r/2^n$$

where r *has one of the values* $0, 1, \quad \cdot, 2^n$. *(The symbol* $\{i, j\}$ *is defined in section 5)*

SKETCH OF PROOF This theorem is proved by investigating how the hyperplanes

$$\tfrac{1}{2}(x_i + x_j) = 0 \qquad (1 \leq i \leq j \leq n),$$

intersect the n-dimensional order statistic space for the particular population considered It is found that each relation of the form

$$\tfrac{1}{2}(x_i + x_j)\ \{i, j\}\ 0, \qquad (1 \leq i \leq j \geq n)$$

defines a region of the n dimensional order statistic space which consists of a certain number r of n-dimensional "basic" cells each of which has an n-dimensional, "volume" equal to $(\frac{1}{2})^n$ A detailed proof of this theorem is given in [5]

Next a method will be developed whereby confidence coefficient values can be determined for any *one-sided* confidence interval of the form

$$\max\ [\text{certain subset of } \tfrac{1}{2}(x_i + x_j),\ 1 \leq i \leq j \leq n] < \phi \tag{7}$$

The confidence coefficient of any one sided confidence interval of the form $\min\ [\ ] > \phi$ can be obtained by symmetry The confidence coefficient of any other one-sided confidence interval of the form (6) can be found by expressing the value of

$$Pr\ [\tfrac{1}{2}(x_i + x_j)\ \{i, j\}\ \phi]$$

as a sum of terms of the form $Pr\{\max\ [\ ] < \phi\}$ or as a sum of terms of the form $Pr\{\min\ [\ ] > \phi\}$. That this is always possible for one-sided confidence intervals of the form (6) is shown by direct application of the results of page 17 of [6].

It is not difficult to show that any one-sided confidence interval of the form (7) can be expressed in the form

$$\max\ [x(n - k),\ \tfrac{1}{2}[x(n - k + 1) + x(n - m_k - k + 1)], \quad \cdot,$$
$$\tfrac{1}{2}[x(n) + x(n - m_1)]] < \phi,$$

where

$$x(i) = x_i, \qquad (i = 1, \cdot \quad , n),$$

and $m_1, \cdots, m_k$ are k integers such that

$$n \geq m_1 > m_2 > \cdots > m_k > 0.$$

This is done by choosing $k, m_1, \cdots, m_k$ so that the two confidence intervals are equivalent.

Thus it is sufficient to prove the following theorem:

THEOREM 4. *Let $x(1), \cdots, x(n)$ represent the ordered values of n independent observations drawn from populations satisfying conditions (A). Choose a set of k integers $m_1, \cdots, m_k$ such that*

$$n \geq m_1 > m_2 > \cdots > m_k > 0$$

Then the one-sided confidence interval

$$\max \{x(n-k), \tfrac{1}{2}[x(n-k+1) + x(n-m_k-k+1)], \cdots, \tfrac{1}{2}[x(n) + x(n-m_1)]\} < \phi, \tag{8}$$

where a term of the form $\frac{1}{2}[x(n-h+1) + x(n-m_h-h+1)]$, $(h = 1, \cdots, k)$, is to be deleted if[2] $n - m_h - h + 1 = 0$, has the confidence coefficient

$$2^{-n}\left[1 + m_1 + \sum_{i_1=1}^{m_2}(m_1 - i_1) + \sum_{i_2=1}^{m_3}\sum_{i_1=1}^{m_1-i_2}(m_1 - i_1 - i_2) + \cdots + \sum_{i_{k-1}=1}^{m_k}\sum_{i_{k-2}=1}^{m_{k-1}-i_{k-1}}\cdots\sum_{i_1=1}^{m_1-i_2-\cdots-i_{k-1}}(m_1 - i_1 - \cdots - i_{k-1})\right] \tag{9}$$

SKETCH OF PROOF. It is sufficient to consider the case in which the n observations are a sample from the uniform population with range $-\frac{1}{2}$ to $\frac{1}{2}$ (then $\phi < 0$).

Let us consider the region of the n-dimensional order statistic space defined by (8). This region can be considered as an intersection of n-dimensional regions each of which is completely defined by a certain region in an x_i, x_j plane $(1 \leq i < j \leq n)$. Also the n-dimensional "volume" of this region equals the value of the confidence coefficient of the confidence interval (8).

By Theorem 3, the intersection region of (8) consists of a certain number of "basic" cells, each of n-dimensional "volume" $(\frac{1}{2})^n$. Theorem 4 is proved by developing a method for finding the number of "basic" cells in this intersection region on the basis of the corresponding regions in the x_i, x_j planes. It is found that the intersection region consists of

$$1 + m_1 + \cdots + \sum_{i_{k-1}=1}^{m_k}\cdots\sum_{i_1=1}^{m_2-i_2-\cdots-i_{k-1}}(m_1 - i_1 - \cdots - i_{k-1})$$

"basic" cells. A detailed derivation of this expression is given in [5].

Now consider some examples of the application of Theorem 4. Let $n = 11$, $m_1 = 11, m_2 = 5, m_3 = 2$. Then, by Theorem 4, the one-sided confidence interval

$$\max [x_3, \tfrac{1}{2}(x_0 + x_7), \tfrac{1}{2}(x_{10} + x_5)] < \phi$$

[2] For the trivial case in which $k = n$ the value of (9) is unity.

has a confidence coefficient equal to $103/2^{11}$ If $n = 12$ instead of 11, the confidence coefficient would be $103, 2^{12}$ while the confidence interval becomes

$$\max [x_2, \tfrac{1}{2}(x_{10} + x_3), \tfrac{1}{2}(x_{11} + x_3), \tfrac{1}{2}(x_{12} + x_4)] < \phi$$

As another example, let $n = 11$ and consider the confidence interval

$$\text{Max } [x_3, \tfrac{1}{2}(x_2 + x_7), \tfrac{1}{2}(x_{10} + x_6), \tfrac{1}{2}(x_{11} + x_4)] < \phi$$

Here $l = 3$ and comparison with (8) shows that this confidence interval satisfied Theorem 1 with $m_1 = 7$, $m_2 = 5$, $m_3 = 2$ Thus it has a confidence coefficient equal to $51/2^{11}$

Theorem 3 shows that each one-sided confidence interval developed on the basis of Theorem 2 has a confidence coefficient of the form $r/2^n$, $(0 \leq r \leq 2^n)$ The question arises as to whether the one sided confidence intervals defined by Theorem 1 have confidence coefficients which attain each of the values $1/2^n$, $2/2^n$, , $(2^n - 1)/2^n$ That this is not the case is proved as follows. The totality of different confidence intervals of the form (8) is equal to $2^n - 1$ This is shown by counting how many ways the integers m_1, , m_k can be selected subject to the conditions $n \geq m_1 > m_2 > > m_k > 0$ It is easily seen that there are $\binom{n}{k}$ possible ways Summing over the possible values of k yields $2^n - 1$ This figure is increased to 2^n if the confidence interval $x_n < \phi$ is also included Examination of (9) shows, however, that two different selections of m_1, m_2, etc, will result in the same value of (9) for more than one case. Thus the one sided confidence intervals of Theorem 1 do not have confidence coefficients which attain each of the values $1/2^n$, , $(2^n - 1)/2^n$

Although the class of one-sided confidence intervals defined by Theorem 4 do not have confidence coefficients which attain each of the values $1/2^n, 2/2^n, \cdots, (2^n - 1)/2^n$, they do have another property which is important from a practical point of view If a certain confidence coefficient can be obtained for a particular value of n, then this confidence coefficient can also be obtained for all greater values of n This result is a consequence of the following theorem

THEOREM 5 *Let $x(1), \cdots, x(n)$ be the ordered values of n independent observations drawn from populations satisfying conditions (A) Then if a confidence interval of the form (8) has the confidence coefficient ϵ for a certain value n_0 of n, it is always possible to obtain another confidence interval of the form* (8), *which has the confidence coefficient ϵ for the value $n_0 + 1$*

PROOF. Let $m_1, \cdots, m_k$ be the integers corresponding to the given confidence interval of form (8) These integers satisfy the condition

$$n_0 \geq m_1 > m_2 > \cdots > m_k > 0$$

Let n_0 be replaced by $n_0 + 1$ and consider the new set of integers $(m_1 + 1)$, $(m_2 + 1), \cdots, (m_k + 1)$, 1. Evidently

$$n_0 + 1 \geq m_1 + 1 > > m_k + 1 > 1 > 0.$$

Hence these integers can be used to define a confidence interval of the form (8). Also it is easily verified that

$$1 + (m_1 + 1) + \sum_{i_1=1}^{m_1+1} (m_1 + 1 - i_1) + \cdots + \sum_{i_{k-1}=1}^{m_{k-1}+1} \cdots \sum_{i_1=1}^{m_1+1-i_2-\cdots-i_{k-1}} (m_1 + 1 - i_1 - \cdots - i_{k-1}) + \sum_{i_k=1}^{1} \sum_{i_{k-1}=1}^{m_{k-1}-i_k} \cdots \sum_{i_1=1}^{m_1+1-i_2-\cdots-i_k} (m_1 - 1 - i_1 - \cdots - i_k) = 2\left[1 + m_1 + \sum_{i_1=1}^{m_1} (m_1 - i_1) + \cdots + \sum_{i_{k-1}=1}^{m_k} \sum_{i_1=1}^{m_2-i_2-\cdots-i_{k-1}} (m_1 - i_1 - \cdots - i_{k-1})\right].$$

Thus the new confidence interval has the same confidence coefficient as the given confidence interval

From symmetry considerations, the one sided confidence interval

$$\min\{x(k+1), \tfrac{1}{2}[x(k) + x(m_k + k)], \cdots, \tfrac{1}{2}[x(1) + x(m_1 + 1)]\} > \phi,$$

where a term of the form $\frac{1}{2}[x(h) + x(m_h + h)]$, $(h = 1, \cdots, k)$, is to be deleted if $m_h + h = n + 1$, has the same confidence coefficient as the one sided confidence interval (8), i e. its confidence coefficient is given by (9)

7 Efficiency of some tests based on conditions (A). Let us consider the case in which the n observations used for a test are a sample from a normal population with unknown variance The purpose of this section is to investigate the efficiency of some tests based on conditions (A) for this special case

The method used to obtain efficiencies is outlined in section 3 Only one-sided and symmetrical tests are considered. For this purpose it is sufficient to limit investigations to one sided tests of $\phi < \phi_0$

If the subset of $\frac{1}{2}(x_i + x_j)$, $(1 \leq i \leq j \leq n)$, chosen for a test is not of one of the forms

(a) x_1

(b) $\frac{1}{2}(x_i + x_j)$, $(i < j)$,

(c) x_j, $\frac{1}{2}(x_i + x_k)$, $(i < j < k)$,

the determination of power function values requires a numerical double or higher order integration Such numerical integrations are extremely lengthy For this reason only one-sided significance tests based on subsets of the forms (a) (c) will be investigated

Let the normal population have variance σ^2 and consider one sided tests of $\phi < \phi_0$ based on subsets of the form (a) Then

$$\text{Power Function} = Pr\,(x_r < \phi_0)$$

$$= Pr\left(\frac{x_r - \phi}{\sigma} < \frac{\phi_0 - \phi}{\sigma}\right) = \sum_{s=1}^{n} \frac{n}{s!\,(n-s)!}\,(N(\delta))^s(1 - N(\delta))^{n-s}$$

where

$$\delta = (\phi_0 - \phi)/\sigma, \qquad N(\delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\delta} e^{-\frac{1}{2}y^2}\,dy$$

The power function values listed for the test $x_r < \phi_0$ in Table 2 were computed from the above expression. The corresponding values for the t-test were computed from the normal approximation given in [2].

For subsets of forms (b) and (c) the expression for the power function is more complicated and will not be either derived or stated here. For any particular case, however, a simple analysis will yield an expression for the power function which requires only a first order numerical integration. General expressions for the power functions when the subsets are of the forms (b) and (c) are stated and derived in [5].

Table 2 contains power function values and efficiencies for several tests based on subsets of the forms (b) and (c). The power function values were computed by approximate integration (Simpson's rule, etc.). The t-test power function values were obtained by using the normal approximation. The power efficiencies listed in Table 1 for tests which do not appear in Table 2 were computed in [5], where a table of power function values is also given.

Examination of Table 2 shows that many of the tests formed from subsets of types (b) and (c) are very efficient for small values of n. The efficiency appears to decrease as n increases. Also the efficiency of a test depends strongly on the subset of $\frac{1}{2}(x_i + x_j)$, $(1 \leq i \leq j \leq n)$, used to form the test. For example, set $n = 10$. The test

$$Accept\ \phi < \phi_0\ if\ \max\,[x_9\,,\ \tfrac{1}{2}(x_1 + x_{10})] < \phi_0$$

has a significance level of approximately .01 but an efficiency of only 82%. However the test

$$Accept\ \phi < \phi_0\ if\ \max[x_8\,,\ \tfrac{1}{2}(x_8 + x_{10})] < \phi_0$$

also has a significance level of approximately .01 but an efficiency of 96.5%.

An approximate set of rules for picking subsets which result in efficient tests of $\phi < \phi_0$ is suggested by the results of Table 2. Let $x(i_1), \cdots, x(i_r)$ be the order statistics which make up the elements of the particular subset of $\frac{1}{2}(x_i + x_j)$, $(1 \leq i \leq j \leq n)$, to be used for the test. The approximate rules are

1. Use the maximum of the values of the elements of the subset.
2. Choose $i_1, \cdots, i_r$ so that $\max(i_1, \cdots, i_r) = n$ and $\min(i_1, \cdots, i_r)$ is as large as possible subject to the restriction that the test is to have a significance level of a specified order of magnitude.

Symmetry considerations furnish the corresponding set of rules for obtaining efficient tests of $\phi < \phi_0$.

Other tests at approximately the same significance level, but not based on subsets of the forms (a)–(c) are undoubtedly more efficient than many of the tests considered in Tables 1 and 2 (particularly for the larger values of n). Computational difficulties, however, prevent consideration of more general situations.

8. A general solution.[3] A general method of obtaining one sided tests of $\phi < \phi_0$ and $\phi > \phi_0$, also symmetrical tests of $\phi \neq \phi_0$, on the basis of conditions (A) is the following.

Let $y_1, \ldots, y_n$ be n independent observations drawn from populations satisfying conditions (A). Let

$$z_i = y_i - \phi_0 \qquad (i = 1, \ldots, n)$$

If the null hypothesis of $\phi = \phi_0$ is satisfied, each z_i is an observation from a population satisfying conditions (A) with zero median. Consider the 2^n sets of values obtained by the transformations

$$z_i \to \epsilon(i) z_i, \qquad (i = 1, \ldots, n)$$

where $\epsilon(i)$ is one of the signs $+$ or $-$. Form the mean of each of the 2^n sets of values. Then it is readily seen, from conditions (A), that the probability that $\bar{z}(= \Sigma z_i/n)$ is less than the $(r+1)$th largest of the 2^n means has the value $r/2^n$ when the null hypothesis is true. Similarly the probability that $\bar{z}$ is greater than the $(2^n - r)$th largest of the 2^n means is equal to $r/2^n$ if the null hypothesis of $\phi = \phi_0$ is satisfied. Thus the test

Accept $\phi < \phi_0$ if $\bar{z}$ is less than the $(r+1)$th largest of the 2^n means

is a one-sided test of $\phi < \phi_0$ with significance level equal to $r/2^n$. Likewise the one-sided test

Accept $\phi > \phi_0$ if $\bar{z}$ is greater than the $(2^n - r)$th largest of the 2^n means

has the significance level $r/2^n$. Consequently the symmetrical test

Accept $\phi \neq \phi_0$ if $\bar{z}$ is either less than the $(r+1)$th largest or greater than the $(2^n - r)$th largest of the 2^n means.

has a significance level equal to $2r/2^n$.

The application of any of the above tests requires the computation of the 2^n means and a determination of where $\bar{z}$ falls in the ordering of these means. If $n = 5$, only 32 means need be computed. If $n = 10$, however, 1021 means must be computed. Evidently this test is too cumbersome to apply except for very small values of n.

9. Acknowledgements. The author would like to express his appreciation to Professors S. S. Wilks and John W. Tukey for valuable advice and assistance in

[3] This solution was derived independently by E. J. G. Pitman and the author. The fundamental idea on which the solution is based was presented by R. A. Fisher in [7].

the preparation of this paper, also to Mrs Ruth S Shafer for computational assistance

REFERENCES

[1] John E Walsh, "Applications of some significance tests for the median which are valid under very general conditions," submitted to *Am Stat Assn Jour*

[2] N L Johnson and B L Welch, "Applications of the non central t distribution," *Biometrika*, Vol 31 (1940), p 370

[3] John E Walsh, "On the power function of the sign test for slippage of means," *Annals of Math Stat*, Vol 17 (1946), pp 360-361

[4] H Scheffé and J W Tukey, "Non parametric estimation I Validation of order statistics," *Annals of Math Stat*, Vol 16 (1945), pp 187–192

[5] John E Walsh, "Some significance tests for the median which are valid under very general conditions," unpublished thesis, Princeton University

[6] G Udny Yule and M G Kendall, *An Introduction to the Theory of Statistics*, Griffin and Co, 1947

[7] R A Fisher, *The Design of Experiments*, Oliver and Boyd, 1942

A DIRECT METHOD FOR PRODUCING RANDOM DIGITS IN ANY NUMBER SYSTEM

By H Burke Horton and R Tines Smith III

Interstate Commerce Commission

1. Summary A compounding technique first used to produce random binary digits is generalized and extended to other number systems Formulae for the rate of convergence of probabilities to the desired values are derived The method is extended to the production of random digits with fixed but unequal probabilities Numerical results are presented in summary form together with results of tests applied to a set of random digits produced by the method

2 Introduction In a note [1] by one of the authors a method of producing random digits was presented. The method was based upon a process, designated "compound randomization," used to produce random binary digits, which can be converted to random digits in other number systems by simple methods Despite the ease of converting a random binary series to another system, it is of interest to examine the problem of direct production of random digits in any number system In the course of producing random binary digits with machine tabulating equipment, and while designing an electronic device to produce random binary digits, it was noted that the multiplication process described in the earlier paper was the equivalent of addition modulo 2 of a series of binary digits. This observation laid the basis for generalizing to other number systems [1]

3 Initial conditions and notation. Let us assume that there is available a source of digits, 0, 1, 2, $\cdots$ $(n-1)$, in a number system of base n, where n is a positive integer, $n > 1$ Let p_{rs} represent the probability of obtaining the rth digit in the sth trial Assume that initial conditions can be controlled so that the trials are independent[2] and

$$p_{rs} \geqq \epsilon \tag{3 1}$$

where $0 < \epsilon \leqq 1/n$ is a fixed positive number (It may be noted at this point that conventional "single-stage" methods of producing random numbers are based upon the assumption that $p_{rs} = \epsilon = 1/n$) Let π_{rs} represent the probability of obtaining the rth digit by addition modulo n of the digits obtained in s individual trials In order to express π_{rs} in terms of p_{rs}, consider two sets of matrices whose elements are defined as follows.

---

[1] In acting as referee for [1] Dr George W Brown suggested generalizing to other number systems by addition modulo n.

[2] J E Walsh [2] has considered, in terms of conditional probabilities, the effect of intercorrelation on compound randomization in the binary system

$$(3\,2)\qquad a_s = \begin{Vmatrix} p_{0,s} & p_{n-1,s} & p_{n-2,s} & \cdots & p_{1,s} \\ p_{1,s} & p_{0,s} & p_{n-1,s} & & p_{2,s} \\ p_{2,s} & p_{1,s} & p_{0,s} & & p_{3,s} \\ \cdots & \cdots & \cdots & & \cdots \\ p_{n-1,s} & p_{n-2,s} & p_{n-3,s} & \cdots & p_{0,s} \end{Vmatrix},$$

$$(3\,3)\qquad \alpha_s = \begin{Vmatrix} \pi_{0,s} & \pi_{n-1,s} & \pi_{n-2,s} & & \pi_{1,s} \\ \pi_{1,s} & \pi_{0,s} & \pi_{n-1,s} & \cdots & \pi_{2,s} \\ \pi_{2,s} & \pi_{1,s} & \pi_{0,s} & & \pi_{3,s} \\ \cdots & \cdots & \cdots & & \cdots \\ \pi_{n-1,s} & \pi_{n-2,s} & \pi_{n-3,s} & & \pi_{0,s} \end{Vmatrix}$$

Note that a_s and α_s are Markoff matrices with two additional restrictions (1) there are no zero elements, and (2) column (as well as row) sums are unity. Each $n \times n$ matrix is made up of only n distinct elements, namely, the n different probabilities associated with the sth trial for a_s, or the n different probabilities associated with the sum of s trials for α_s.

4. Relation of π_{rs} to p_{rs}. Assuming independent trials, we have the following relationships

$$(4\,1)\qquad \begin{aligned} \alpha_1 &= a_1, \\ \alpha_2 &= a_2 \quad a_1 = a_2 \quad a_1, \\ \alpha_3 &= a_3 \quad a_2 \quad a_1 = a_3 \quad a_2 \quad a_1, \\ &\cdots \cdots \cdots \\ \alpha_k &= a_k \quad a_{k-1} \quad \cdots \quad a_1 = \prod_{s=1}^{k} a_s \end{aligned}$$

Thus, since any row (or any column) of α_k is a permutation of the π_{rk}, by (4.1) the π_{rk} are expressed in terms of the individual probabilities, p_{rs}.

5. Convergence of π_{rs} to $1/n$. (5.01) THEOREM[3] $\operatorname{Lim}_{k\to\infty} \pi_{rk} = 1/n$.

PROOF. Let ρ_s denote the range of the elements of α_s. Each element of α_s is a weighted mean of the n distinct elements of α_{s-1}. The n distinct elements of a_s are used as weights in the averaging process. Now the range of a set of weighted means (weights > 0) of a set of values must be less than the range of the values themselves, unless both ranges are zero. Therefore, since the weights, $p_{rs} > 0$ by condition (3.1),

$$(5.02)\qquad \rho_s < \rho_{s-1}, \text{ for } \rho_{s-1} \neq 0, \text{ or in the special case } \rho_{s-1} = 0,\ \rho_s = 0.$$

Also, since $\sum_{r=0}^{n-1} \pi_{rs} = 1$,

[3] While this article was awaiting publication, J. Wolfowitz independently proved theorem (5.01).

$$1/n - \rho_s \leqq \pi_{rs} \leqq 1/n + \rho_s \tag{5.03}$$

In order to show that $\lim_{s\to\infty} \rho_s = 0$, and to derive formulae for the rate of convergence of π_{rs} to the limiting value, $1/n$, let w_i represent the ordered p_{rs} for any given s w_1 = the smallest p_{rs}, w_n = the largest of the p_{rs}. In a similar manner let x_i represent the ordered $\pi_{r\,s-1}$. The following inequalities for the maximum and minimum π_{rs} can be set down immediately.

$$\max_r \pi_{rs} \leqq w_n \cdot x_n + w_{n-1} x_{n-1} + \cdots + w_1 \cdot x_1, \tag{5.04}$$

$$\min_r \pi_{rs} \geqq w_n x_1 + w_{n-1} \cdot x_2 + \cdots + w_1 x_n \tag{5.05}$$

And since $\rho_s = \max_r \pi_{rs} - \min_r \pi_{rs}$,

$$\rho_s \leqq w_n(x_n - x_1) + w_{n-1}(x_{n-1} - x_2) + \cdots + w_2(x_2 - x_{n-1}) + w_1(x_1 - x_n). \tag{5.06}$$

For n even, let $m = n/2 + 1$, then by regrouping terms,

$$\begin{aligned}\rho_s \leqq\ & (w_n - w_1)(x_n - x_1) + (w_{n-1} - w_2)(x_{n-1} - x_2) + \cdots \\ & + (w_m - w_{m-1})(x_m - x_{m-1})\end{aligned} \tag{5.07}$$

Noting that $\rho_{s-1} = (x_n - x_1) \geqq (x_{n-1} - x_2) \geqq \cdots \geqq (x_m - x_{m-1})$, the following substitutions can be made

$$\rho_s \leqq (w_n - w_1)\rho_{s-1} + (w_{n-1} - w_2)\rho_{s-1} + \cdots + (w_m - w_{m-1})\rho_{s-1}. \tag{5.08}$$

For compactness, this may be written,

$$\rho_s \leqq \left[\sum_{i=m}^{n} w_i - \sum_{i=1}^{m-1} w_i\right] \cdot \rho_{s-1}. \tag{5.09}$$

Similarly for n odd, let $m = (n+1)/2$; proceeding in the same manner as above, the median term vanishes, yielding as a final result,

$$\rho_s \leqq \left[\sum_{i=m+1}^{n} w_i - \sum_{i=1}^{m-1} w_i\right] \cdot \rho_{s-1}. \tag{5.10}$$

For simplicity denote the expression in brackets by δ_s, then

$$\rho_s \leqq \delta_s \cdot \rho_{s-1}, \tag{5.11}$$

where for n even, δ_s represents the sum of the largest $n/2$ of the p_{rs} minus the sum of the smallest $n/2$ of the p_{rs}, and for n odd δ_s represents the sum of the largest $(n-1)/2$ of the p_{rs} minus the sum of the smallest $(n-1)/2$ of the p_{rs}. Continuing the process developed above, we find that

$$\rho_s \leqq \delta_s \quad \delta_{s-1} \cdot \rho_{s-2}; \tag{5.12}$$

$$\rho_s \leqq \delta_s \quad \delta_{s-1} \cdot \delta_{s-2} \cdot \cdots \quad \delta_2 \cdot \rho_1. \tag{5.13}$$

Since $\delta_1 \geqq \rho_1$, the following simple inequality holds

$$\rho_k \leqq \prod_{i=1}^{k} \delta_i, \tag{5 14}$$

Now $\delta_i \leqq 1 - n\epsilon$, by condition (3 1) and the definition of δ_i Therefore,

$$\lim_{k\to\infty} \rho_k \leq \lim_{k\to\infty} \prod_{i=1}^{k} \delta_i \leq \lim_{k\to\infty} (1 - n\epsilon)^k = 0, \tag{5 15}$$

and (5 01) is proven In the special case of constant probabilities from trial to trial, $\delta_i = \delta_0$, a constant,and (5 14) becomes

$$\rho_k \leqq (\delta_0)^k \tag{5 16}$$

Since the mean π_{ri} is $1/n$, we have the following useful inequalities·

$$1/n - \prod_{i=1}^{k} \delta_i \leqq \pi_{rk} \leqq 1/n + \prod_{i=1}^{k} \delta_i, \tag{5 17}$$

in the case of varying probabilities, and

$$1/n - (\delta_0)^k \leqq \pi_{rk} \leqq 1/n + (\delta_0)^k, \tag{5 18}$$

in the case of constant probabilities If δ_i is not known in each trial, an upper bound, δ_b, may be estimated on the basis of knowledge (including statistical tests) of the digit generating process Then the following inequality will hold

$$1/n - (\delta_b)^k \leqq \pi_{rk} \leqq 1/n + (\delta_b)^k, \tag{5 19}$$

where $\delta_b \leqq (1 - n\epsilon)$

It is worthy of note that inequalities (5 11) and (5 15) become equalities if $n = 2$ (binary system), thus,

$$\rho_k = \prod_{i=1}^{k} \delta_i = \prod_{i=1}^{k} |p_i - q_i| = \prod_{i=1}^{k} |2p_i - 1|, \tag{5 14b}$$

$$\rho_k = (\delta_0)^k = |p - q|^k = |2p - 1|^k. \tag{5 15b}$$

These results were obtained by different methods in [1]

8. Discussion of results. Certain facts are implicit in the foregoing analysis, but are worthy of mention in passing. The compounding process may consist of addition modulo n of digits taken from a number of digit-producing machines If any machine, h, is perfect, i e., $p_{rh} = 1/n$ for all r, each element of the probability matrix a_h will be equal to $1/n$, and $\rho_h = 0$ Consequently, each element of α_s, $s \geqq h$, will be equal to $1/n$ by (5 17) and the special case of (5 02) Thus any combination which contains a perfect machine is perfect. This is equivalent to a restatement of Von Mises' [3] requirement that the sum of a random set and any other set must itself be a random set Furthermore, by (5 02) the results taken from any machine, no matter how nearly perfect, can be improved

by combining with the results of another machine, no matter how biased the latter may be In the limiting case, $p_{rs} = 1$ (or 0), the probabilities of the various digits are merely interchanged

7. Production of random numbers with fixed but unequal probabilities. The principles presented above can be adapted to the production of random numbers with unequal probabilities as follows: Assume that a set of random digits, 0, 1, 2, $(n-1)$, is required in a number system of base n, with probabilities q_0, q_1, q_2, q_{n-1}, $\sum_{i=0}^{n-1} q_i = 1$, where each q_i is a proper rational fraction which may be written as the quotient of two positive integers, $q_i = \frac{u_i}{v_i}$. Choose m as the basis of a new number system, where m is the least common multiple of the v_i,

$$q_i = \frac{u_i}{v_i} = \frac{mu_i/v_i}{m}. \tag{7.1}$$

A set of random digits, 0, 1, 2, $(m-1)$, in a number system of base m may be generated by the process described above, or a set of such digits may be constructed by entering an existing table of random digits, base n, and interpreting appropriate numerical quantities, base n, as digit symbols, base m Since $\frac{mu_i}{v_i}$ is an integer, groups of digits, mu_0, mu_1, mu_{n-1}, in the m system may be coded as digits, 0, 1, 2, $(n-1)$, in the n system An upper bound for the maximum bias of q_i will be $\frac{mu_i}{v_i}\rho_k$, where ρ_k is the range of π_{rk} in the n system Thus, by increasing k, the bias of q_i can be made smaller than any preassigned quantity

8 Convergence under more general conditions. Convergence of π_{rs} to $1/n$ occurs under a variety of conditions less restrictive than (3.1)

(8.1) THEOREM *In the case of independent trials, a necessary and sufficient condition that* $\lim_{s\to\infty} \pi_{rs} = 1/n$ *is that* $\left\{\begin{matrix}\pi_{0l}\\ \pi_{1l}\end{matrix}\right\} \geqq \epsilon$, *where* ϵ *is a fixed positive number, arbitrarily small, and* l *is a fixed positive integer, arbitrarily large.* It is obvious that (8.1) is a necessary condition for convergence. To prove that it is a sufficient condition, consider the following:

(8.2) LEMMA. *If* $\left\{\begin{matrix}p_{0s}\\ p_{1s}\end{matrix}\right\} \geqq \eta$, *where* η *is a fixed positive number, arbitrarily small, then* $\lim_{s\to\infty} \pi_{rs} = 1/n$

PROOF Take a fixed integer, h, $h \geqq n-1$. Now any digit, r, can be obtained in at least one way, i.e., as the sum of r *ones* and $(h-r)$ *zeros*. Therefore,

$$\pi_{rh} \geqq \tau, \qquad \text{where } \tau = \eta^h \tag{8.3}$$

We now regard h trials as a single trial of a complex machine. Let u represent the number of such complex trials. Let π'_{ru} represent the probability of obtaining the rth digit as the result of addition modulo n of u complex trials. Then,

$$\lim_{u\to\infty} \pi'_{ru} = \lim_{u\to\infty} \pi_{r,(uh)} = 1/n, \tag{8.4}$$

by (5.01). Now $s = uh + j$, $0 \leqq j < h$, (j an integer), or $uh \leqq uh + j < (u+1)h$. The j simple trials cannot increase the maximum bias, by (5.02); consequently,

$$\lim_{u\to\infty} \pi_{r,(uh+j)} \qquad \lim_{(uh+j)\to\infty} \pi_{r,(uh+j)} = 1/n \tag{8.5}$$

Since there is a one-to-one correspondence between the elements of $\{s\}$ and $\{uh + j\}$,

$$\lim_{s\to\infty} \pi_{rs} = 1/n \tag{8.6}$$

By a natural extension of the lemma, we may regard t trials as a single complex trial. Theorem (8.1) thus assumes the form of (8.2).

9. Numerical results in various number systems. More efficient convergence formulae can be devised to meet special conditions. Those presented in (5) have the advantages of simplicity and generality. To test the efficiency of (5.15) several numerical examples, based upon unusual hypothetical probabilities, were worked by matrix multiplication as in (4.1). In these problems $p_{rs} = p_r$, a constant, from trial to trial. A tabular comparison of the ranges, computed by (4.1), and the upper bounds, determined by (5.15), is presented in Table 1 for $k = 10$.

10. Preparation and tests of a set of random digits. Since an unlimited number of valid tests for randomness may be devised, it is obvious that any finite set of digits cannot meet all such tests. As a matter of fact a truly random process should yield sets which *fail* to meet some proportion of the tests, the fraction being determined by the level of significance adopted in testing. No finite set of digits can be considered random, the tests for randomness are really applied to determine the character of the generating process. However, the concept of "locally random" sets as developed by Kendall and Smith [4] is useful, and some of their tests are used below as evidence that a set of numbers produced by compound randomization is likely to be locally random.

A non-random set of 10,000 decimal digits having the relative frequencies indicated in the starred line of Table 1 was punched in cards and tabulated. Totals were taken for each ten cards and the amount in the unit's position of the counter was cut in a summary card, thereby producing a set of 1,000 digits. The frequencies of digits in the derived set are compared with those of the generating set in Table 2. The frequencies of the derived set are in accord with the hypothesis of equal probabilities.

TABLE 1

Comparison of computed range and formula for maximum base, $k = 10$
Hypothetical numerical examples, constant probabilities from trial to trial

Number base	Probability in an individual trial												p_0	$[R_0]$	k_0
	p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}			
2	800	200	—	—	—	—	—	—	—	—	—	—	[illegible]	[illegible]	[illegible]
3	500	300	200	—	—	—	—	—	—	—	—	—	[illegible]	[illegible]	[illegible]
3	070	020	010	—	—	—	—	—	—	—	—	—	[illegible]	[illegible]	[illegible]
3	400	300	300	—	—	—	—	—	—	—	—	—	[illegible]	[illegible]	[illegible]
4	200	100	400	300	—	—	—	—	—	—	—		[illegible]	[illegible]	[illegible]
5	050	200	400	020	330	—	—	—	—	—	—		[illegible]	[illegible]	[illegible]
6	080	240	360	020	200	100	—	—	—	—	—	—	[illegible]	[illegible]	[illegible]
7	300	020	240	050	130	170	090	—	—	—	—	—	[illegible]	[illegible]	[illegible]
8	200	050	060	180	100	090	150	110	—	—	—	—	[illegible]	[illegible]	[illegible]
9	030	080	150	060	140	000	100	050	210	—	—	—	[illegible]	[illegible]	[illegible]
10	050	150	200	050	050	120	080	020	180	100	—		[illegible]	[illegible]	[illegible]
10	010	020	030	010	050	000	070	060	090	550	—	—	[illegible]	[illegible]	[illegible]
10	110	110	110	110	110	110	110	110	110	010	—		[illegible]	[illegible]	[illegible]
10	150	150	150	150	150	050	050	050	050	050	—		[illegible]	[illegible]	[illegible]
10*	014	171	164	184	023	095	017	205	089	008	—		[illegible]	[illegible]	[illegible]
12	010	070	120	100	050	020	090	010	080	110	000	100	[illegible]	[illegible]	[illegible]

* This badly biased set of probabilities was used to produce the set of random decimal digits tested in the next section

TABLE 2

Digit	0	1	2	3	4	5	6	7	8	9
Generating set	014	171	164	184	023	095	017	205	089	008
Derived set	088	112	086	105	113	102	101	098	097	098

Frequency test (derived set) $\chi^2 = 7.0$ $P = .63$

TABLE 3

sth digit	$(s+1)$th digit									
	0	1	2	3	4	5	6	7	8	9
0	11	8	7	7	5	7	12	12	11	8
1	10	13	15	9	11	11	11	8	10	11
2	11	10	7	10	10	7	9	9	7	9
3	9	10	3	14	12	17	9	8	11	12
4	6	12	10	10	19	6	10	11	13	7
5	9	17	11	14	10	6	5	15	6	9
6	6	14	9	9	14	10	15	8	6	10
7	13	10	9	9	8	11	7	12	7	12
8	7	8	8	12	9	11	14	9	10	10

In the serial test adjacent pairs of digits are tabulated The distribution of these pairs in the derived set appears in Table 3 This test indicates that adjacent digits are independent

TABLE 4

Gap test

Digit		Length of gap				χ^2	P
		0-1	2 4	5-7	8 and over		
		Frequencies					
0	Observed	16	18	11	42	1 25	75
	Expected	16.53	19.10	13.02	37 45		
1	Observed	27	27	21	36	5 44	15
	Expected	21 09	24.37	17.76	47 78		
2	Observed	16	17	10	42	1 90	.60
	Expected	16 15	18 60	13 60	36 59		
3	Observed	19	26	18	41	90	92
	Expected	19 70	22.83	16.04	44 77		
4	Observed	31	17	20	44	7.39	.06
	Expected	21 28	21.59	17 92	48 21		
5	Observed	15	21	15	50	2 04	.57
	Expected	19 19	22.17	16 16	43 48		
6	Observed	27	25	12	36	5 95	12
	Expected	19 00	21 95	16 00	43.05		
7	Observed	20	19	16	42	40	.93
	Expected	18 13	21 29	15 52	41 70		
8	Observed	14	19	21	42	3 27	35
	Expected	18 21	21 07	15.36	41 32		
9	Observed	18	18	21	40	2 53	.48
	Expected	18.43	21 29	15 52	41 76		

The gap test is based upon the distribution of lengths of intervals between given digits A comparison of the number of gaps of specified lengths and the expected number in each case is presented in Table 4 The results of this test

are also in accord with the assumption of local randomness. Noting the badly biased probabilities of the initial set of digits, the results of these tests demonstrate the effectiveness of the compound randomization process.

The use of tabulating equipment for producing random decimal digits by addition modulo 10 is relatively fast and simple. The authors have just completed production of a set of 105,000 digits in less than two days' tabulating time. 75,000 cards, representing approximately 3 months' receipts of a current railroad waybill study, were used to generate the digits, 11 non-correlated columns being added simultaneously. A chain of length 10 was used, although the nature of the initial data was such that a shorter length would probably have given satisfactory results. The derived set is now recorded on 1500 cards, 70 digits per card. Preliminary tests for local randomness confirm the random nature of the generating process. Upon completion of the tests this set will be reproduced in tabular form.

REFERENCES

[1] H. B. Horton, "A method for obtaining random numbers," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 81–85.

[2] J. E. Walsh, "Concerning compound randomization in the binary system," unpublished manuscript, *Project RAND, Douglas Aircraft Co.*, Santa Monica, California.

[3] R. von Mises, *Probability, Statistics and Truth*, The Macmillan Co., New York, 1939.

[4] M. G. Kendall and B. B. Smith, "Randomness and random sampling numbers," *Roy. Stat. Soc. Jour.*, Vol. 101 (1938), pp. 147–166.

[5] M. G. Kendall and B. B. Smith, "Second paper on random sampling numbers," *Supp. to Roy. Stat. Soc. Jour.*, Vol. 6 (1939), pp. 51–61.

[6] G. U. Yule, "A test of Tippett's random sampling numbers," *Roy. Stat. Soc. Jour.*, Vol. 101 (1938), pp. 167–172.

[7] C. W. Vickery, "On drawing a random sample from a set of punched cards," *Supp. to Roy. Stat. Soc. Jour.*, Vol. 6 (1939), pp. 62–66.

ON A MATCHING PROBLEM ARISING IN GENETICS

By Howard Levene

Columbia University

1. Summary. A statistic useful for detecting deviations from the Hardy-Weinberg equilibrium in population genetics is discussed. Both exact and asymptotic distributions are given and a special case where there is misclassification is discussed. The distribution obtained also arises from a certain card matching problem.

2. Introduction. A system of multiple alleles behaves as follows under Mendelian inheritance. There are r distinct forms of alleles, $a_1, \cdots, a_r$, of a given gene. A given individual contains two genes and can be represented as a_i/a_j. If $i = j$ the individual is called a homozygote, if $i \neq j$ it is called a heterozygote. The representation a_i/a_j is called the genotype. In reproduction each gamete produced by an a_i/a_j individual contains one gene which has a probability 1/2 of being a_i and 1/2 of being a_j. In fertilization a paternal and a maternal gamete fuse to form a new individual which contains two genes, giving the well-known Mendelian ratios. We now consider a large *random breeding* population of N individuals. This will contain $2N$ genes, of which the proportion q_i will be of type $a_i (i = 1, \cdots, r, \Sigma q_i = 1)$. The probability that a random individual from the next generation will be a_i/a_j is $q_i^2 (i = j)$ or $2q_iq_j (i \neq j)$, which are known as the Hardy-Weinberg equilibrium probabilities. The statistical problem arose in testing (by means of a sample of n individuals) the hypothesis that this Hardy-Weinberg ratio holds against the alternative hypothesis that disturbing forces decrease the number of homozygotes. The actual data has been discussed elsewhere [1].

3. The sample distribution of number of homozygotes. We shall assume throughout this paper that N is so large that random fluctuations in the population proportions from generation to generation can be ignored. Let $x_{ij} (i \leq j = 1, \cdots, r)$ be the number of a_i/a_j individuals in the sample, and let $y_i = x_{ii} + \Sigma_{j=1}^{r} x_{ij}$ be the number of a_i genes in the sample. We have $\Sigma\Sigma x_{ij} = n$ and $\Sigma y_i = 2n$. Let $h = \Sigma x_{ii}$ be the number of homozygotes, and $z = n - h$ be the number of heterozygotes in the sample. The probability of the observed sample is

$$P = \frac{n!}{\prod_{j \leq i} x_{ij}!} \prod_{i=1}^{r} (q_i^{2x_{ii}}) \prod_{i<j} (2q_iq_j)^{x_{ij}} = \frac{n!2^z}{\prod_{j \leq i} x_{ij}!} \prod_{i=1}^{r} q_i^{y_i}. \tag{1}$$

Since the q_i are unknown we use the conditional probability when $y_1, \cdots, y_r$ are held constant. Whenever we use the word "conditional" hereafter, this condition will be understood. The conditional probability is

$$K' \frac{n!2^s}{\prod_{i \le j} x_{ij}!}, \quad \text{where}$$

$$\frac{1}{K'} = \sum' \frac{n!2^s}{\prod_{i \le j} x_{ij}!}, \tag{2}$$

where the summation Σ' is over all non-negative integral values of the x_{ij} subject to the condition

$$x_{ii} + \Sigma_j x_{ij} = y_i \qquad (i = 1, \cdots, r)$$

Consider

$$\left(\sum_1^r t_i\right)^{2n} = (\sum t_i^2 + 2\sum_{i<j} t_i t_j)^n \tag{3}$$

$$= \sum^* \frac{n!2^s}{\prod_{i \le j} x_{ij}!} \prod_{i=1}^r t_i^{x_{ii}+\sum_{j>i} x_{ij}}, \tag{4}$$

where the summation Σ^* is over all non-negative values of the x_{ij} subject to the condition $\Sigma_{i \le j} x_{ij} = n$. Evidently $1/K'$ is the coefficient of $\Pi t_i^{y_i}$ in (1), but this must equal the coefficient of this term in the left member of (3), and thus $1/K' = (2n)!/\Pi y_i!$. Hence the conditional probability of the observed sample is

$$P = \frac{n!\prod y_i!}{(2n)!} \cdot \frac{2^s}{\prod_{i \le j} x_{ij}!} \tag{5}$$

For any function $u(x_{11}, \cdots, x_{1r}, \cdots, x_{rr})$ we will now let $E(u)$ and $\sigma^2(u)$ denote the *conditional* mean and variance of u for fixed y_i, and will refer to them simply as the mean and variance. We first obtain the sth factorial moment of x_{ii}, that is $E(x_{ii}^{(s)})$, where $x^{(s)} = x(x-1)\cdots(x-s+1)$. Consider

$$\sum' \frac{2^s n!}{\prod_{j \le k} x_{jk}!} x_{ii}^{(s)} = n^{(s)} \sum' \frac{2^s (n-s)!}{\prod_{j \le k} x'_{jk}!}, \tag{6}$$

where $x'_{jk} = x_{jk}$ except that $x'_{ii} = x_{ii} - s$, and Σ' has the same meaning as in (2). The right member of (6) is evaluated exactly as before, giving

$$E(x_{ii}^{(s)}) = \frac{n^{(s)}(y_i)^{(2s)}}{(2n)^{(2s)}}. \tag{7}$$

From this expression we obtain

$$E(x_{ii}) = \frac{y_i(y_i-1)}{4n-2} = n\beta_i^2 + O(1), \tag{8}$$

and

$$\sigma^2(v_{ii}) = \frac{n^{(4)}y_i^{(4)}}{(2n)^{(4)}} + \frac{ny_i^{(2)}}{(2n)^{(2)}} - \left[\frac{ny_i^{(2)}}{(2n)^{(2)}}\right]^2 = nf_i^2(1-f_i)^2 + O(1), \tag{9}$$

where $f_i = y_i/2n$ is the sample estimate of q_i. Similarly

$$E(x_{ii}^{(4)}x_{jj}^{(0)}) = \frac{n^{(2+0)}y_i^{(2a)}y_j^{(2b)}}{(2n)^{(2a+2b)}} \tag{10}$$

giving

$$\sigma(v_{ii}, v_{jj}) = \frac{n^{(2)}y_i^{(2)}y_j^{(2)}}{(2n)^{(4)}} - \frac{y_i^{(2)}y_j^{(2)}}{4(2n-1)^2} \tag{11}$$

Other moments can be similarly evaluated, in particular $E(v_{ij}) = y_iy_j/(2n-1)$.

4. Asymptotic distribution of number of homozygotes. From (8), (9), and (11) we may easily obtain

$$E(h) = \Sigma E(v_{ii}) = (C-2n)/(2n-2), \tag{12}$$

$$\sigma^2(h) = \Sigma\sigma^2(v_{ii}) + 2\sum\sum_{i<j}\sigma(v_{ii}, v_{jj}) \tag{13}$$

$$= \frac{1}{4n^2}\left\{C(n+2) + C^2\left(\frac{2n+5}{8n^2}\right) - D\left(\frac{n+2}{n}\right)\right\} - \tfrac{1}{2} + O\left(\frac{1}{n}\right), \tag{14}$$

where $C = \Sigma y_i^2$ and $D = \Sigma y_i^4$. The formula (14) is a close approximation to (13) and is easily computed. From (5) by means similar to those classically used to prove asymptotic normality of the binomial distribution we can prove asymptotic normality of the conditional distribution of h, more precisely, if $n \to \infty$ and $y_i/n \to$ constant ($i = 1, \quad , r$), then

$$\text{Prob}\left\{\frac{h - E(h)}{\sigma(h)} \le t\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}\,dx. \tag{15}$$

5. Effect of misclassification. There is a further complication in the particular case reported in [1]. All individuals of genotype a_i/a_i are correctly classified, but an individual of genotype a_i/a_j ($i \neq j$) has a known probability $p/2$ of being classified a_i/a_i and an equal probability of being classified a_j/a_j. As a result, the observed proportion of homozygotes is a biased estimate of the proportion in the population. Let h, v_{ii}, , y_i denote the true sample values, and let h', v_{ii}', , y_i' denote the recorded sample values. Then $h^* = h' - e$, where $e = (n-h')\,p/(1-p)$, will give an unbiased estimate, i.e. $E(h^*) = E(h)$. In order to use h^* we must have its (conditional) variance. Since $h^* = np/(1-p) + h'/(1-p)$,

$$\sigma_{h^*}^2 = [1/(1-p)]^2\sigma_{h'}^2.$$

Let $h - h' = \epsilon$, then for large fixed $(n-h)$, ϵ is approximately normally distributed with mean $(n-h)p$ and variance

$$(n - h)p(1 - p) = [n - E(h)]p(1 - p)[1 + O_p(1/\sqrt{n})]$$

Neglecting the remainder term in this variance, ϵ and h have a joint normal distribution with parameters that are easily calculated. We thus have

$$\sigma_{h'}^2 = \sigma_h^2 + \sigma_\epsilon^2 + 2\sigma(h, \epsilon), \quad \text{or} \quad \sigma_{h'}^2 = [n - E(h)]p(1 - p) + (1 - p)^2\sigma_h^2,$$

giving

$$\sigma_{h'}^2 = \sigma_h^2 + [n - E(h)]p/(1 - p). \tag{16}$$

In [1] $\sigma_{h'}^2$ was given as $\sigma_h^2 + \epsilon$ for the sake of simplicity. This would tend to be smaller than (16), but only negligibly so. Strictly speaking the calculation of $E(h)$ and σ_h^2 from (12) and (14) requires a knowledge of the true y_i, but the observed y_i' are unbiased estimates of the y_i and their use should cause no serious trouble.

6. Combinatorial statement of the problem. This problem can also be expressed as one of card matching as follows. A deck contains $2n$ cards of r different suits, with y_i cards of the ith suit ($i = 1, \quad , r$). We draw n pairs of cards at random without replacement, exhausting the deck. What is the distribution of h, the number of twins (pairs in which both members are of the same suit). If $z = n - h$, the probability of exactly h twins is given by (7), and in the limit h is normally distributed with mean given by (12) and variance given by (14). The card matching problem does not involve the notion of conditional probability. By introducing variables u_α equal to one if the αth pair is a twin and zero otherwise, the moments of h can also be obtained without using generating functions.

REFERENCE

[1] Theodosius Dobzhansky and Howard Levene, "Genetics of natural populations XVII. Proof of operation of natural selection in wild populations of *Drosophila pseudoobscura*," *Genetics*, Vol. 33 (1948), pp. 537-547.

A MULTIPLE DECISION PROCEDURE FOR CERTAIN PROBLEMS IN THE ANALYSIS OF VARIANCE

By Edward Paulson

University of Washington

1 Introduction. In this paper we will discuss a certain type of problem which arises in many applications of the analysis of variance. We suppose that we are given K varieties, and are required to investigate the differences among them on the basis of the observed yields from a given experimental design, such as a set of randomized blocks or a latin square. The classical procedure [1] for dealing with this problem has been to test the null hypothesis that the K varieties are all equal by computing the ratio of the mean sum of squares between varieties to the residual mean sum of squares, and rejecting the null hypothesis whenever this ratio exceeded the critical value corresponding to the level of significance used. However, the standard discussions of this procedure seem to be quite vague on the question of what action should be taken after the null hypothesis has been rejected.

In a number of problems, the practical situation seems to be such that instead of testing the null hypothesis that the varieties do not differ, what is really required is a statistical rule or "decision function" which on the basis of the observed yields will classify the K varieties into a "superior" group and an "inferior" group. If the superior group consists of more than one variety, the next appropriate action will of course depend on the particular problem at hand. In some situations the varieties in the superior group might then be subject to further selection on the basis of some secondary characteristic, or additional observations might be taken to discriminate between the members of the superior group, after discarding the varieties in the inferior group. However, if all varieties happen to be classified in one group, the group will be labelled "neutral" and this result is to be interpreted as implying that the varieties are homogeneous.

In this formulation, the problem is now of a multiple decision type, it is necessary to decide on the basis of a sample which one out of the $2^K - 1$ possible decisions (or classifications) to select. We will suggest a solution which seems quite reasonable on an intuitive basis, but it is still an open question whether this solution is an optimum one.

2. A special case. In this section we will discuss the problem under the assumption that the variance σ^2 of a single observation is known *a priori*. This is a rather restrictive assumption, but it can be considered as approximately satisfied when the number of degrees of freedom available for estimating the variance is large, which will often be the case. The minor modifications necessary to secure exact results for the small sample case when σ is unknown are

discussed in section 3 We also assume that the experimental design has been so selected that there will be the same number (r) of observations on each of the K varieties

Now let $x_{i\alpha}$ = the αth observation on the ith variety ($i = 1, 2, \cdots, K$; $\alpha = 1, 2, \cdots, r$), let $\bar{x}_i = \sum_{\alpha=1}^{r} x_{i\alpha}/r$, put $m_i = E(\bar{x}_i)$ where E stands for expected value, and take λ to be a given positive constant. The conventional assumption is made that all the observations are normally and independently distributed with the same variance σ^2. Denote by $\bar{x}_M$ the maximum of the K mean values $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_K$. The rule for dividing the varieties into superior and inferior groups is the following: the superior group is to consist of all varieties whose corresponding mean values fall in the interval $[\bar{x}_M - \lambda\sigma/\sqrt{r}, \bar{x}_M]$ and the remaining varieties constitute the inferior group. (As mentioned earlier, if all the varieties fall into one group, this group is labelled 'neutral' and the varieties are considered homogeneous.)

This rule completely determines the classification as soon as λ is determined. For a given sample size, we might select λ by considering the relative importance of different types of incorrect classifications. If H denotes the error of misclassifying the varieties when in fact they are all equal, and G denotes the error of misclassifying the varieties when they actually are unequal, then it is obvious that the greater the value of λ, the smaller the probability of an error of type H, but the greater the probability of an error of type G. Therefore for a given value of r it is necessary to adopt some sort of compromise in selecting λ.

For a given value of λ we will now derive explicit formulas for $P(H)$, the probability of not classifying all the varieties in one group when $m_1 = m_2 = \cdots = m_K$, and for $P(G_1)$ the probability that as a result of the experiment there will not be a superior group consisting only of the Kth variety when $m_1 = m_2 = \cdots = m_{K-1} = m$ and $m_K = m + \Delta (\Delta > 0)$. G_1 was selected because it appeared to be the particular kind of type G error most likely to be useful in applications. Also $P(G_1)$ may be regarded as the least upper bound of the probability of misclassifying the varieties when one variety is superior to any of the others by an amount at least equal to Δ. Now if we denote by $W = (\bar{x}_M - \bar{x}_{\min})$ the difference between the maximum and minimum values of the set $\{\bar{x}_i\}$ ($i = 1, 2, \cdots, K$), then it is obvious that

$$1 - P(H) = P\left\{W < \frac{\lambda\sigma}{\sqrt{r}}\right\}. \tag{2.1}$$

The right hand side of (2.1) is equivalent to the probability that the range of a sample of K independent observations from a normal distribution with unit variance be less than λ; this probability has already been tabulated by Pearson and Hartley [2]. From these tables it is a routine matter to find $P(H)$ corresponding to a given value of λ, and conversely. To evaluate $P(G_1)$, we have

$$1 - P(G_1) = P\left\{\bar{x}_i < \bar{x}_K - \frac{\lambda\sigma}{\sqrt{r}} \quad \text{for each } i \quad (i = 1, 2, \cdots, K-1)\right\}.$$

By evaluating the probability of this event for a fixed value of $\bar{x}_K$ and then integrating out with respect to $\bar{x}_K$, it is a simple matter to verify that

$$P(G_1) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2/2)} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y + (\Delta/\sigma)\sqrt{r} - \lambda} e^{-(t^2/2)}\, dt \right]^{K-1} dy. \tag{2.2}$$

In some applications, it may be desirable to have an explicit expression for the probability that the superior group will consist of the Kth variety and not more than s inferior varieties when $m_1 = m_2 = \cdots = m_{K-1} = m$ and $m_K = m + \Delta$. If we denote this probability by $1 - P_s^*$ it is not difficult to show that

$$1 - P_s^* = \sum_{\alpha=0}^{s} \binom{K-1}{\alpha} [T_{1\alpha} + \alpha T_{2\alpha}], \quad \text{where}$$

$$T_{1\alpha} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2/2)} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y + (\Delta/\sigma)\sqrt{r} - \lambda} e^{-(t^2/2)}\, dt \right]^{K-\alpha-1}$$

$$\cdot \left[\frac{1}{\sqrt{2\pi}} \int_{y + (\Delta/\sigma)\sqrt{r} - \lambda}^{y + (\Delta/\sigma)\sqrt{r}} e^{-(t^2/2)}\, dt \right]^{\alpha} dy, \quad \text{and} \tag{2.3}$$

$$T_{2\alpha} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2/2)} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y - \lambda} e^{-(t^2/2)}\, dt \right]^{K-\alpha-1}$$

$$\cdot \left[\frac{1}{\sqrt{2\pi}} \int_{y - \lambda}^{y} e^{-(t^2/2)}\, dt \right]^{\alpha - 1} \left[\frac{1}{\sqrt{2\pi}} \int_{y - (\Delta/\sigma)\sqrt{r} - \lambda}^{y - (\Delta/\sigma)\sqrt{r}} e^{-(t^2/2)}\, dt \right] dy.$$

3. General case. We now briefly discuss the exact treatment of the problem when σ is unknown. The notation of section 2 will be used, but in addition denote by s^2 an estimate of σ^2 resulting from the given experimental design which is based on the residual sum of squares with n degrees of freedom. It is well known that s^2 is independent of the set $\{\bar{x}_i\}$ $(i = 1, 2, \cdots K)$. Now the rule to be used in classifying the varieties into two groups is as follows: the superior group is to consist of all those varieties whose mean values fall in the interval $[\bar{x}_M - \lambda s/\sqrt{r}, \bar{x}_M]$, and the inferior group consists of the remaining varieties.

We now find that:

$$1 - P(H) = P\{W \leq \lambda s/\sqrt{r}\}. \tag{3.1}$$

The right hand side of (3.1) depends only on the distribution of the 'studentized' range and has also been tabulated by Pearson and Hartley [3] although the tabulation is considerably less complete than that of the range in [2]. It is also easy to verify that the expression for $P(G_1)$ now becomes

$$P(G_1) = 1 - \frac{n^{n/2}}{\sqrt{2\pi}\, 2^{(n-2)/2} \Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} \int_{-\infty}^{\infty} w^{n-1} e^{-(nw^2 + y^2)/2}$$

$$\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y + (\Delta/\sigma)\sqrt{r} - \lambda w} e^{-(t^2/2)}\, dt \right]^{K-1} dy\, dw \tag{3.2}$$

with a similar modification for P_s^*.

4. Remarks. Any application of the ideas suggested here would be greatly facilitated if tables of $P(G_1)$ were made available. If this were done, it would be possible to decide in advance of an experiment how large r should be in order to have a fixed control over both types H and G_1 errors. It is obvious that further research both along theoretical and applied lines is needed. In conclusion, the writer would like to thank Professor Albert Bowker for several helpful suggestions.

REFERENCES

[1] R. A Fisher, *Statistical Methods for Research Workers*, Chapters 7, 8.

[2] E. S Pearson and H. O Hartley, "Tables of the probability integral of the range in samples from a normal population," *Biometrika*, Vol 32 (1941–42), pp 301–310.

[3] E. S Pearson and H. O Hartley, "Tables of the probability integral of the studentized range," *Biometrika*, Vol. 33 (1943), pp. 89–99.

A MODIFIED EXTREME VALUE PROBLEM

By Benjamin Epstein[1]

Coal Research Laboratory, Carnegie Institute of Technology

1. Introduction and summary. Consider the following problem.

Particles are distributed over unit areas in such a way that the number of particles to be found in such areas is a random variable following the law of Poisson, with ν equal to the expected number of particles per unit area. Furthermore, the particles themselves are assumed to vary in magnitude according to a size distribution specified (independently of the particular unit area chosen) by a d.f. $F(x)$ defined over some interval $a \leq x \leq b$, with $F(a) = 0$ and $F(b) = 1$. The problem is to find the distribution of the smallest, largest, or more generally the nth smallest or nth largest particle in randomly chosen unit areas.

The problem as stated is not completely specified. To specify the distribution of smallest or largest particles in a unit area one must give a rule for dealing with those areas which contain no particles at all. More generally, in the case of the distribution of the nth smallest or nth largest particle, one must give a rule for dealing with those areas which contain $(n - 1)$ or fewer particles There are at least two possible alternatives. One alternative is to omit none of the areas from consideration by setting up the following rule: if no particles are found in a given unit area then this area will be considered as one for which the smallest size particle is $x = b$ and for which the largest size particle is $x = a$. More generally, if $(n - 1)$ or fewer particles are found in a given unit area then this area will be considered as one for which the nth smallest size particle is $x = b$ and for which the nth largest size particle is $x = a$. A second alternative is to restrict attention to those areas which contain at least one particle (in the case of the distribution of smallest or largest values) or at least n particles (in the case of the distribution of the nth smallest or nth largest particle). In other words, this means finding the relevant conditional distribution.

From the point of view of the application of the theory of extreme values to fracture problems, there are some situations where the first model and other situations where the second model is the more appropriate in describing the phenomenon under investigation. In this paper section 2 will be devoted to a derivation of the distributions associated with the first alternative; in section 3 the conditional distributions will be described briefly.

2. The distributions under the first alternative. In this section we shall be concerned with the first alternative. To find the distribution of the nth smallest particle in unit areas, we first observe (the verification is left to the

[1] Present address, Department of Mathematics, Wayne University, Detroit, Michigan.

reader) that under the hypotheses of section 1, the number of particles having size $\leq x$ in a unit area is distributed according to the law of Poisson, with expected number equal to $\nu F(x)$. Next we note that the probability that the nth smallest particle in a unit area exceeds x in size is equal to the probability of finding exactly 0, or exactly 1, or exactly 2, $\cdots$, or exactly $(n-1)$ particles of size $\leq x$ in that area. Therefore $G_n(x)$, the probability that the nth smallest size particle in a unit area is $\leq x$, is given by

$$G_n(x) = 1 - \sum_{j=0}^{n-1} e^{-\nu F(x)} \frac{(\nu F(x))^j}{j!}, \qquad x < b; \tag{1}$$

$$= 1, \qquad x \geq b,$$

where we have assigned to the size $x = b$ the probability $\sum_{j=0}^{n-1} e^{-\nu}(\nu^j/j)!$ which is just equal to the probability of finding fewer than n particles in a unit area.

If the d.f. $F(x)$ has a derivative $f(x)$ for all x lying in $a \leq x \leq b$, then $G_n(x)$ has a derivative for any value of $x \neq b$. Therefore the probability density for the nth smallest size particle is, for any $x \neq b$, given by the function $g_n(x)$ where

$$g_n(x) = e^{-\nu F(x)} \frac{(\nu F(x))^{n-1}}{(n-1)!} \nu f(x), \qquad a \leq x < b; \tag{2}$$

$$= 0, \qquad x < a, \qquad x > b.$$

A finite probability $\sum_{j=0}^{n-1} e^{-\nu} \frac{\nu^j}{j!}$ is assigned to $x = b$.

If one makes the transformation $y = \nu F(x)$ (for a similar transformation in extreme value theory see [1, page 371]), then (1), and (2) become

$$G_n^*(y) = 1 - \sum_{j=0}^{n-1} e^{-\nu} \frac{y^j}{j!}, \qquad y < \nu; \tag{1'}$$

$$= 1, \qquad y \geq \nu,$$

and

$$g_n^*(y) = \frac{e^{-\nu} y^{n-1}}{(n-1)!}, \qquad 0 \leq y < \nu; \tag{2'}$$

$$= 0, \qquad y < 0, \qquad y > \nu.$$

A finite probability $\sum_{j=0}^{n-1} e^{-\nu} \frac{\nu^j}{j!}$ is assigned to $y = \nu$.

The distribution of the smallest size particle in a randomly chosen area is found by letting $n = 1$ in equation 1.

In a similar way one can find the distribution of the nth largest particle in a randomly chosen unit area. $H_n(x)$, the probability that the nth largest size particle in a unit area is $\leq x$, is given by

$$
\begin{aligned}
H_n(x) &= 0, \qquad x < a; \\
&= \sum_{j=0}^{n-1} e^{-\nu(1-F(x))} \frac{[\nu(1 - F(x))]^j}{j!}, \qquad x \geq a,
\end{aligned} \tag{3}
$$

where we have assigned to the size $x = a$ the probability $\sum_{j=0}^{n-1} e^{-\nu} \frac{\nu^j}{j!}$.

If, as before, $F(x)$ is assumed to have a derivative $f(x)$ for all x lying in $a \leq x \leq b$, then the probability density for the nth largest size particle is, for any $x \neq a$, given by the function $h_n(x)$ where

$$
\begin{aligned}
h_n(x) &= e^{-\nu(1-F(x))} \frac{[\nu(1 - F(x)]^{n-1}}{(n-1)!} \nu f(x), \qquad a < x \leq b; \\
&= 0, \qquad x < a, \qquad x > b.
\end{aligned} \tag{4}
$$

A finite probability $\sum_{j=0}^{n-1} e^{-\nu} \frac{\nu^j}{j!}$ is assigned to $x = a$.

If one makes the transformation $z = \nu[1 - F(x)]$, then (3) and (4) become

$$
\begin{aligned}
H_n^*(z) &= 1 - \sum_{j=0}^{n-1} e^{-z} \frac{z^j}{j!}, \qquad z < \nu; \\
&= 1, \qquad z \geq \nu,
\end{aligned} \tag{3'}
$$

and

$$
\begin{aligned}
h_n^*(z) &= \frac{e^{-z} z^{n-1}}{(n-1)!}, \qquad 0 \leq z < \nu; \\
&= 0, \qquad z < 0, \qquad z > \nu,
\end{aligned} \tag{4'}
$$

with a finite probability $\sum_{j=0}^{n-1} e^{-\nu} \frac{\nu^j}{j!}$ assigned to $z = \nu$.

The distribution of the largest size particle in a randomly chosen unit area is found by letting $n = 1$ in equation 3.

3. Conditional distributions of the extreme values. The appropriate conditional distributions for the problem under consideration can be written down readily. The step function component which occurred in section 2 is no longer present since we restrict our attention only to those areas which contain at least n particles (in the general case of the distribution of nth smallest or nth largest size particles).

$G_n^c(x)$, the d.f. of the nth smallest particle in a unit area chosen at random from the class of areas containing at least n particles, is given by

$$
\begin{aligned}
G_n^c(x) &= 0, \qquad x < a; \\
&= \frac{1 - \sum_{j=0}^{n-1} e^{-\nu F(x)} (\nu F(x))^j / j!}{1 - \sum_{j=0}^{n-1} e^{-\nu} \nu^j / j!}, \qquad a \leq x \leq b; \\
&= 1, \qquad x > b.
\end{aligned} \tag{5}
$$

Similarly $H_n^c(x)$, the d f. of the nth largest particle in a unit area chosen at random from the class of areas containing at least n particles, is given by

$$
\begin{aligned}
H_n^c(x) &= 0, \qquad x < a, \\
&= \frac{\sum_{j=0}^{n-1} e^{-\nu[1-F(x)]}[\nu(1-F(x))]^j/j! - \sum_{j=0}^{n-1} e^{-\nu}\nu^j/j!}{1 - \sum_{j=0}^{n-1} e^{-\nu}\nu^j/j!}, \qquad a \leq x \leq b; \\
&= 1, \qquad x > b.
\end{aligned} \tag{6}
$$

4. General remarks and an application. It is interesting to note that the assumptions of section 1 lead to distribution functions in section 2 which are precisely the same as the asymptotic distributions of smallest, largest, or nth smallest, or nth largest values in samples of fixed size $N (N \to \infty)$ (see e g. [1, p. 371]). In the problem treated in this paper, ν, the expected number of particles in a unit area, plays the role of N in the fixed sample size case, with the important difference that the distributions in the present paper are exact and not merely asymptotic.

The results of this paper have a direct bearing on certain aspects of fracture problems [2] and in particular on the dielectric breakdown of capacitors [3]. In the latter problem there appears to be ample justification for assuming that the breakdown voltage is influenced to a considerable degree by the presence of flaws known in the technical literature as conducting particles. These particles are spread individually and collectively at random throughout the area of the capacitor and, depending on their size, create a local weakening of the capacitor by reducing the nominal insulation thickness in the neighborhood of flaws. The voltage required to break down the capacitor is equal to that required to break it down at that spot where the greatest penetration has taken place.

In the dielectric problem the statistical distribution of largest values appropriate to the problem is given by (3) with $n = 1$, and the size distribution of conducting particles follows a law of the form $f(x) = \lambda e^{-\lambda x}$, $x > 0$. This is a situation where all the capacitors under test are part of the sample (since all must be tested to destruction) and those which happen to contain no defects (an event with probability $e^{-\nu}$) act as if the largest particle size is equal to $a = 0$. $e^{-\nu}$ simply represents the expected fraction of capacitors which have strength equal to the theoretical strength of the insulation.

The conditional distributions of section 3 would be more appropriate in the following sort of practical situation. Suppose that surface flaws spread at random on glass rods are known to reduce greatly the strength of the rods. Suppose that in a given sample of glass rods one takes out by some method of inspection those specimens which have no flaws. Then the strength distribution of the remaining specimens is a conditional distribution since each specimen must contain at least one flaw to be eligible as a member of the sample.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] B. Epstein, "Statistical aspects of fracture problems," *J. Applied Phys.*, Vol. 19 (1948), pp. 140-147.
[3] B. Epstein and H. Brooks, "The theory of extreme values and its implications in the study of the dielectric strength of paper capacitors," *J. Applied Physics*, Vol. 19 (1948), pp. 544-550.

ON DISTINCT HYPOTHESES

By Agnes Berger and Abraham Wald

Columbia University

1. Introduction. The following problem was suggested to one of the authors by Professor Neyman:

Let $X = (X_1, X_4, \cdots, X_n)$ be a chance vector and let h denote any simple hypothesis specifying its distribution Let H_i be the composite hypothesis that some element h of a set of simple hypotheses $\{h\}_i$, $(i = 0, 1)$, is true, and assume that H_0 and H_1 are known to be exhaustive. Let h_i denote an element of $\{h\}_i$ $(i = 0, 1)$.

For any region W of the sample space S, let $P(W \mid h)$ be the probability that the sample point falls in W when h is true.

We shall call H_0 and H_1 *distinct*, if a region W exists for which

$$P(W \mid h_0) \neq P(W \mid h_1), \qquad \begin{matrix} \text{for all } h_0 \in \{h\}_0 \\ \text{and all } h_1 \in \{h\}_1 . \end{matrix}$$

The problem is to establish necessary and sufficient conditions for two composite hypotheses H_0 and H_1 to be distinct.

For any critical region W for testing H_0 against H_1, let $\gamma(W \mid h)$ be the probability of a wrong decision when h is true, i.e.

$$\gamma(W \mid h) = \begin{cases} P(W \mid h) & \text{for } \ h \in H_0 \\ 1 - P(W \mid h) & \text{for } \ h \in H_1 . \end{cases}$$

Suppose now that H_0 and H_1 are not distinct. Then to any W a pair h_0', h_1' exist such that

$$P(W \mid h_0') = P(W \mid h_1'),$$

thus

$$\gamma(W \mid h_0') = 1 - \gamma(W \mid h_1'),$$

and therefore

$$\text{l.u.b } \gamma(W \mid h) \geq \tfrac{1}{2} \text{ for any } W. \tag{1.1}$$

This property of non-distinct hypotheses leads us to investigate the conditions under which 2 hypotheses allow a test where the maximum probability of a wrong decision is $< \frac{1}{2}$.

The result, in turn, will enable us to state, for an important class of hypotheses a necessary and sufficient condition for 2 composite hypotheses to be distinct.

2. A lemma. We shall now prove the following lemma:

Lemma 2 1. *Assume that X has a density function $p(x)$ and let $H_i = h_i$ be the simple hypothesis that $p(x) = p_i(x)$, $(i = 0, 1)$. Assume that the set R of x's*

satisfying $p_0(x) \neq p_1(x)$ *has a positive measure. Then there exists a region* W *such that* $\gamma(W \mid p_i) < \frac{1}{2}$, $i = 0, 1$.

PROOF: Let R_0 be defined by $p_0 = p_1$, R_1 by $p_0 < p_1$, R_2 by $p_0 > p_1$. Since $\int_S p_i(x)\,dx = 1$ and $p_i(x) \geq 0$, $(i = 0, 1)$, R_1 and R_2 are of positive measure. Let

$$\phi(x) = \begin{cases} p_1 \text{ in } R_1 \\ p_0 \text{ in } R_2 \\ p_1 = p_0 \text{ in } R_3 . \end{cases}$$

Then $\int_S \phi(x)\,dx > 1$ and either

a) $\int_{R_1+R_3} p_1\,dx > \frac{1}{2}$ or b) $\int_{R_2} p_0\,dx > \frac{1}{2}$

or both. Assume first a).

Let $R_3 \subset R_1 + R_0$ and such that $\int_{R_3} p_1\,dx = \frac{1}{2}$, but $\int_{R_3} p_0\,dx < \frac{1}{2}$. This can be done by including into R_3 a part of R_1 of non-zero measure. Let $R_4 \subset R_1 + R_0 - R_3$ and such that $0 < \int_{R_4} p_1\,dx < \frac{1}{2} - \int_{R_3} p_0\,dx$. Then

$\int_{R_4} p_0\,dx < \int_{R_4} p_1 dx < \frac{1}{2} - \int_{R_3} p_0\,dx$, thus $\int_{R_3+R_4} p_0\,dx < \frac{1}{2}$ but $\int_{R_3+R_4} p_1\,dx > \frac{1}{2}$.

Assume now b).

Let $R_5 \subset R_2$ and such that $\int_{R_5} p_0\,dx = \frac{1}{2}$. Then $\int_{R_5} p_1\,dx < \frac{1}{2}$. Let $R_6 \subset R_2 - R_5$ and such that $0 < \int_{R_6} p_0\,dx < \frac{1}{2} - \int_{R_5} p_1\,dx$. Then $\int_{R_5+R_6} p_0\,dx > \frac{1}{2}$ and $\int_{R_5+R_6} p_1\,dx < \frac{1}{2}$.

Thus in case a) $W = R_3 + R_4$, and in case b) $W = S - R_5 - R_6$ is a critical region for which $\gamma(W \mid p_i) < \frac{1}{2}$ $(i = 0, 1)$. This proves the lemma.

3. The main theorem. Assume now X to have a density function $p(x, \mid \theta)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ is an unknown parameter point. Let ω_0 and ω_1 be two disjoint, bounded and closed subsets of the k-dimensional θ — space. Let $\Omega = \omega_0 + \omega_1$ and suppose that θ is known to belong to Ω, which therefore will be called the parameter space. Let H_i be the hypothesis that the true parameter point is an element of ω_i, $(i = 0, 1)$.

We shall consider the problem of testing H_0 against H_1. Clearly, $P(W \mid h)$ can now be written as $P(W \mid \theta)$ and $\gamma(W \mid h)$ as $\gamma(W \mid \theta)$.

We shall make the following assumptions concerning $p(x \mid \theta)$:

Assumption 1. $p(x \mid \theta)$ *is continuous in* θ. This is of course always fulfilled if Ω consists only of a finite number of points.

Assumption 2. For any bounded domain M *of the sample space we have*

$$\int_M [\operatorname{Max}_\theta p(x \mid \theta)]\, dx < \infty.$$

It follows from Assumptions 1. and 2. that

$$\lim_{r=\infty} \int_{S-S_r} p(x \mid \theta)\, dx = 0 \tag{3.1}$$

uniformly in θ where S_r is the sphere in the sample space with center at the origin and radius r.

In what follows, whenever we shall speak of cumulative distribution function $g(\theta)$ in the k-dimensional parameter space, we shall always mean a cumulative distribution function satisfying the condition

$$\int_\Omega dg\ (\theta) = 1.$$

For any c.d.f. $g(\theta)$ let W_g denote a critical region which contains any sample point x satisfying the inequality

$$\int_{\omega_1} p(x \mid \theta)\, dg(\theta) > \int_{\omega_0} p(x \mid \theta)\, dg(\theta),$$

and does not contain a sample point x for which

$$\int_{\omega_1} p(x \mid \theta)\, dg(\theta) < \int_{\omega_0} p(x \mid \theta)\, dg(\theta).$$

It can easily be verified that W_g minimizes the average risk

$$\int_\Omega \gamma(W \mid \theta)\, dg(\theta), \text{ i.e., } \int_\Omega W_g \mid \theta)\, dg(\theta) = \operatorname{Min}_W \int_\Omega \gamma(W \mid \theta)\, dg(\theta). \tag{3.2}$$

Let Ω_i $(i = 0, 1)$ be the class of all density functions $p(x) = \int_\Omega p(x \mid \theta)\, dg_i(\theta)$ where $g_i(\theta)$ is subject to the condition

$$\int_{\omega_i} dh_i(\theta) = 1.$$

Two density functions $p(x)$ and $q(x)$ are said to be equal if $p(x) \neq q(x)$ holds only in a set of measure zero.

It follows from (3.1) and Assumptions 1. and 2. that $\gamma(W \mid \theta)$ is a continuous function of θ. Let $\gamma(W)$ denote the maximum of $\gamma(W \mid \theta)$ with respect to θ. We shall prove the following theorem:

THEOREM 3.1 *A necessary and sufficient condition for the existence of a region* W *such that* $\gamma(W) < \frac{1}{2}$ *is that the classes* Ω_0 *and* Ω_1 *be disjoint.*

PROOF. Suppose that Ω_0 and Ω_1 are not disjoint. Then there exist two distribution functions $g_0(\theta)$ and $g_1(\theta)$ such that

$$\int_{\omega_0} dg_0(\theta) = \int_{\omega_1} dg_1(\theta) = 1$$

and

$$\int_{\omega_0} p(x \mid \theta)\, dg_0(\theta) = \int_{\omega_1} p(x \mid \theta)\, dg_1(\theta)$$

(except perhaps for points x in a set of measure 0).

Let $g(\theta) = \frac{1}{2} g_0(\theta) + \frac{1}{2} g_1(\theta)$. Clearly, $\gamma(W) \geq \int_\Omega \gamma(W \mid \theta)\, dg(\theta) = \frac{1}{2}$ for any W. This proves the necessity of our condition.

We shall now assume that Ω_0 and Ω_1 are disjoint. First we shall show that the results of [1] can be applied. On pages 297-8 of [1] there are seven conditions listed for the sequential case. For the non-sequential case (the one considered here) the conditions 6 and 7 drop out and the first five conditions can be reduced to the following conditions:

Condition 1: The weight function $W(\theta, d)$ is bounded

Condition 2: For any θ, the chance vector X admits a density function $p(x \mid \theta)$.

Condition 3: For any sequence $\{\theta_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ there exists a subsequence $\{\theta_{i_j}\}$ $(j = 1, 2, \cdots)$ and a parameter point θ_0 such that

$$\lim_{i \to \infty} p(x \mid \theta_{i_j}) = p(x \mid \theta_0)$$

Condition 4: If $\{\theta_i\}$ $(i = 1, 2, \cdots)$ is a sequence of points and θ_0 a point such that

$$\lim_{i \to \infty} p(x \mid \theta_i) = p(x \mid \theta_0)$$

then,

$$\lim_{i \to \infty} W(\theta_i, d) = W(\theta_0, d)$$

uniformly in d.

Condition 5: The same as our Assumption 2.

In our problem d (the decision of the statistician) can take only two values: acceptance or rejection of H_0. Condition 1 is evidently fulfilled, since $W(\theta, d) = 0$ if a correct decision is made, and $= 1$ if a wrong decision is made. Clearly, Conditions 2 5 are also fulfilled in our problem.

A distribution $g(\theta)$ is said to be least favorable, if it maximizes the minimum average risk, i.e., if it maximizes $\int_\Omega \gamma(W \mid \theta)\, dg(\theta)$ with respect to g.

It follows from Theorems 4.1 and 4.4 of [1] that there exists a least favorable distribution.

Let $g^*(\theta)$ be a least favorable distribution. Then, as has been shown in [1] there exists a W_{g^*} such that

$$\max_{\theta} \gamma(W_{g^*} \mid \theta) = \int_{\Omega} \gamma(W_{g^*} \mid \theta)\, dg^*(\theta). \tag{3.3}$$

Thus, our theorem is proved if we can show that

$$\int_{\Omega} \gamma(W_{g^*} \mid \theta)\, dg^*(\theta) < \tfrac{1}{2}. \tag{3.4}$$

Let H_0^* be the hypothesis that the true density is given by

$$p_0(x) = \frac{\int_{\omega_0} p(x \mid \theta)\, dg^*(\theta)}{\int_{\omega_0} dg^*(\theta)},$$

and H_1^* the hypothesis that the true density is given by

$$p_1(x) = \frac{\int_{\omega_1} p(x \mid \theta)\, dg^*(\theta)}{\int_{\omega_1} dg^*(\theta)}.$$

Since Ω_0 and Ω_1 are disjoint, $p_0(x)$ and $p_1(x)$ are different density functions. Hence, according to Lemma 2.1, there exists a critical region W^* for testing H_0^* such that $\alpha^* < \frac{1}{2}$ and $\beta^* < \frac{1}{2}$, where α^* is the probability of type I error, and β^* is the probability of type II error. Clearly,

$$\tfrac{1}{2} > \alpha^* \int_{\omega_0} dg^*(\theta) + \beta^* \int_{\omega_1} dg^*(\theta) = \int_{\Omega} \gamma(W^* \mid \theta)\, dg^*(\theta). \tag{3.5}$$

Hence, our theorem is proved

It follows from (1.1) that if H_0 and H_1 are not distinct, Ω_0 and Ω_1 are not disjoint.

On the other hand, suppose that Ω_0 and Ω_1 are not disjoint and let

$$\int_{\omega_0} p(x \mid \theta)\, dg_0(\theta) = \int_{\omega_1} p(x \mid \theta)\, dg_1(\theta).$$

Then for every W

$$\int_{\omega_0} P(W \mid \theta)\, dg_0(\theta) = \int_{\omega_1} P(W \mid \theta)\, dg_1(\theta). \tag{3.6}$$

Assume now that ω_i is a connected set ($i = 0, 1$). Then, because of the continuity of $P(W \mid \theta)$ there exist 2 functions $\theta_0(W)$, $\theta_1(W)$, $\theta_i(W)$ belonging to $\omega_i (i = 0, 1)$ such that

$$P(W \mid \theta_0(W)) = \int_{\omega_0} P(W \mid \theta)\, dg_0(\theta)$$

and

$$P\{W \mid H_1\} = \int_{\Omega_1} P(W \mid \theta)\, dg_1(\theta)$$

for every W. Hence, for [illegible]

$$P\{W \mid H_1\} = P(W \mid \theta_1(W))$$

for every W. Thus, [illegible] the following theorem:

THEOREM 3.2. *If* [illegible] $(0, 1)$, *then, under the assumptions of Theorem 3.1,* [illegible] *condition for H_0 and H_1 to be distinct is that the* [illegible] *and W_1* [illegible]

REFERENCE

[1] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. [illegible]

AN APPROXIMATION TO THE SAMPLING VARIANCE OF AN ESTIMATED MAXIMUM VALUE OF GIVEN FREQUENCY BASED ON FIT OF DOUBLY EXPONENTIAL DISTRIBUTION OF MAXIMUM VALUES[1]

By Bradford F. Kimball

N. Y. State Department of Public Service

1. Introduction. Given the doubly exponential distribution of maximum values

$$F(x) = \exp(-e^{-y}), \qquad y = \alpha(x - u), \tag{1}$$

where α and u are unknown parameters, with a prescribed frequency F_0 the "reduced variate" y is fixed, say at $y = y_0$. Thus with

$$F_0 = .99, \qquad y_0 = 4.60015 \cdots .$$

Given a sample of n maximum values x_i, we are interested in the sampling variance of

$$\hat{x} = g(\hat{u}, \hat{\alpha}) = \hat{u} + y_0/\hat{\alpha} \tag{2}$$

due to sampling variations of the estimates $\hat{u}$ and $\hat{\alpha}$.

H Fairfield Smith has recently pointed out to me that the examples of applications of sufficient statistical estimation functions to this problem given in a previous paper (see [1, pp. 307–309]) give too large a range for $\hat{x} = g(\hat{u}, \hat{\alpha})$ because the sample points $(\hat{u}, \hat{\alpha})$ within the confidence region of the constant probability ellipse apply to optimum estimates of $(\hat{u}, \hat{\alpha})$ rather than to that of $g = g(\hat{u}, \hat{\alpha})$. What the problem calls for is the determination of the positions of curves $\bar{g}(u, \alpha)$ and $g(u, \alpha)$ such that the integral of the *pdf* of the estimation functions over all sample values $(\hat{u}, \hat{\alpha})$ which lie between these two curves is equal to the confidence level (taken as .95 in previous paper). Further considerations of this being the shortest interval $\bar{g} - g$, also come into play.

As so often happens in research, the previous analysis, although not giving the final answer, suggests the next step. If we change our parameters to

$$g = g(u, \alpha) = u + y_0/\alpha, \qquad \alpha' = \alpha \tag{3}$$

and are able to carry through the inverse of the maximum likelihood solution for fitting of (1) to n sample values x_i, then we shall be in a position to find the asymptotic marginal distribution of $\sqrt{n}(\hat{g} - g)$, which will give the answer to our problem (see [2]).

The Jacobian of this transformation of parameters is

$$\partial(u, \alpha)/\partial(g, \alpha') = \begin{vmatrix} 1 & y_0^2/\alpha'^2 \\ 0 & 1 \end{vmatrix} = 1,$$

and hence for $\alpha' > 0$ no new singularities are introduced.

[1] This involves a correction of a previous paper [1].

2. The equations of the maximum likelihood solution. For a sample of size n, the *pdf* of the sampling distribution in terms of the old parameters is given by

$$P[u, \alpha, O_n(x_i)] = \alpha^n \exp[-\Sigma e^{-\alpha(x_i-u)}] \exp[-\Sigma\alpha(x_i - u)],$$

and

$$\log P = n \log \alpha - \Sigma e^{-\alpha(x_i-u)} - \alpha\Sigma x_i + n\alpha u;$$
$$= n[\log \alpha - e^{\alpha u}(\Sigma e^{-\alpha x_i}/n) - \alpha\bar{x} + \alpha u].$$

Now change to the new parameters and use the substitutions:

$$z_i = e^{-\alpha x_i}, \qquad \bar{z} = (\Sigma z_i)/n, \qquad z_0 = e^{-\alpha u} = e^{y_0} \cdot e^{-\alpha' g}.$$

Thus

$$\partial z_0/\partial g = -\alpha' z_0, \qquad \partial z_0/\partial \alpha' = -g z_0,$$

and denoting $\log P$ by L we write

$$L = n[\log \alpha' - \bar{z}/z_0 - \alpha'\bar{x} + \alpha' g - y_0].$$

Hence

$$L_g = -n\alpha'[\bar{z}/z_0 - 1]; \tag{4}$$

$$L_{\alpha'} = n[1/\alpha' - \partial(\bar{z}/z_0)/\partial\alpha' - \bar{x} + g]. \tag{5}$$

3. Derivation of expected values needed. Recall that

$$\bar{z}/z_0 = e^{-y_0}\Sigma e^{-\alpha'(x_i-g)}/n = \Sigma e^{-\alpha(x_i-u)}/n.$$

Hence

$$\partial(\bar{z}/z_0)/\partial\alpha' = -e^{-y_0}\Sigma(x_i - g)e^{-\alpha'(x_i-g)}/n, \tag{6}$$
$$\partial(\bar{z}/z_0)/\partial\alpha = -\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n;$$

$$\partial^2(\bar{z}/z_0)/\partial\alpha'^2 = e^{-y_0}\Sigma(x_i - g)^2 e^{-\alpha'(x_i-g)}, \tag{7}$$
$$\partial^2(\bar{z}/z_0)/\partial\alpha^2 = \Sigma(x_i - u)^2 e^{-\alpha(x_i-u)}/n.$$

By investigation of the generating function

$$G(l) = E[\Sigma(z_i/z_0)^{l-1}], \qquad z_i = e^{-\alpha x_i},$$

it can be shown that

$$E[\Sigma e^{-\alpha(x_i-u)}/n] = 1,$$

$$E[\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n] = -(1/\alpha)\Gamma'(2) = -(1/\alpha)(1 - C),$$

where C denotes Euler's constant, .577216 $\cdots$, and

$$E[\Sigma(x_i - u)^2 e^{-\alpha(x_i-u)}/n] = (1/\alpha^2)\Gamma''(2) = (1/\alpha^2)(\pi^2/6 + C^2 - 2C).$$

Hence to find expected values of (6) and (7) we note that

$$-e^{-y_0}\Sigma(x_i - g)e^{-\alpha'(x_i-g)}/n = -\Sigma(x_i - g)e^{-\alpha(x_i-u)}/n;$$
$$= -\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n + (y_0/\alpha)\Sigma e^{-\alpha(x_i-u)}/n,$$

and therefore

$$E[\partial(\bar{z}/z_0)/\partial\alpha'] = E[\partial(\bar{z}/z_0)/\partial\alpha] + (y_0/\alpha)\,E(\bar{z}/z_0). \tag{8}$$

Similar analysis shows that

$$E[\partial^2(\bar{z}/z_0)/\partial\alpha'^2] = E[\partial^2(\bar{z}/z_0)/\partial\alpha^2] + (2y_0/\alpha)E[\partial(\bar{z}/z_0)/\partial\alpha] + (y_0^2/\alpha^2)E[\bar{z}/z_0]. \tag{9}$$

4. The inverse of the maximum likelihood solution. It will first be noted that the maximum likelihood equations (4) and (5) for determining best estimates of g and α' become identical to those for determining best estimates of old parameters u and α, when the transformation of parameters (3) is applied to them. This is easily verified by applying relations developed above.[2]

This means that *the best estimates $\hat{g}$ and $\hat{\alpha}'$ obtained from* (4) *and* (5) *are related to the best estimates of old parameters $\hat{u}$ and $\hat{\alpha}$ by*

$$\hat{g} = \hat{u} + y_0/\hat{\alpha}, \qquad \hat{\alpha}' = \hat{\alpha}. \tag{10}$$

We now proceed to set up the inverse of the maximum likelihood solution. In order to do this we first need the variance-covariance matrix of the direct solution. This is (see [2])

$$\left\| \begin{matrix} E[-L_{gg}] & E[-L_{g\alpha'}] \\ E[-L_{\alpha' g}] & E[-L_{\alpha'\alpha'}] \end{matrix} \right\|.$$

Now

$$L_{gg} = -n\alpha'^2(\bar{z}/z_0), \qquad E[-L_{gg}] = n\alpha'^2,$$
$$L_{g\alpha'} = -n[\bar{z}/z_0 - 1 + \alpha'\partial(\bar{z}/z_0)/\partial\alpha'], \qquad E[L_{g\alpha'}] = n(1 - C + y_0),$$
$$L_{\alpha'\alpha'} = -n[1/\alpha'^2 - \partial^2(\bar{z}/z_0)/\partial\alpha'^2],$$
$$E[-L_{\alpha'\alpha'}] = (n/\alpha'^2)[\pi^2/6 + (1 - C + y_0)^2].$$

Thus the variance-covariance matrix of the estimation functions (4) and (5) is

$$\left\| \begin{matrix} n\alpha'^2 & n(1 - C + y_0) \\ n(1 - C + y_0) & (n/\alpha'^2)[\pi^2/6 + (1 - C + y_0)^2] \end{matrix} \right\|.$$

The asymptotic form of the inverse solution for $\sqrt{n}\,(\hat{g} - g)$ and $\sqrt{n}\,(\hat{\alpha}' - \alpha')$ will have the variance-covariance matrix which is the reciprocal of the above matrix, multiplied by n. The determinant value of the above matrix reduces to $n^2(\pi^2/6)$. Thus the reciprocal matrix, adjusted by multiplying by n, is

[2] See equations (5.2) of [1] and note $+\partial(\bar{z}/z_0)/\partial\alpha$ in second equation of (5.2) should read $-\partial(\bar{z}/z_0)/\partial\alpha$.

$$(11)\qquad \begin{Vmatrix} \alpha'^2[1 + (1 - C + y_0)^2/(\pi^2/6)] & -(1 - C + y_0)/(\pi^2/6) \\ -(1 - C + y_0)/\pi^2/6) & \alpha'^2/(\pi^2/6) \end{Vmatrix}.$$

This gives the solution sought. From the general theory of the maximum likelihood solution (see [2]) the distribution of $[\sqrt{n}(\hat{g} - g), \sqrt{n}(\hat{\alpha}' - \alpha')]$ is asymptotically normal. Hence *the marginal distribution of* $\sqrt{n}(\hat{g} - g)$ *will be asymptotically normal, and for finite* n, *the standard deviation may be approximated by*

$$(12)\qquad \sigma(\hat{g} - g) = [1/(\sqrt{n}\alpha')]\sqrt{1 + (1 - C + y_0)^2/(\pi^2/6)}.$$

Now the correlation coefficient for the asymptotic bivariate normal distribution is seen to be

$$r = -(1 - C + y_0)/\sqrt{\pi^2/6 + (1 - C + y_0)^2}.$$

If α' were known, we should have the standard deviation of $\sqrt{n}(\hat{g} - g)$ reduced by factor $\sqrt{1 - r^2}$. This is found to be equal to the reciprocal of the second factor in the equation (12). Hence we conclude that *if* α' *be known, the standard deviation of* $(\hat{g} - g)$, *for finite* n, *is given approximately by*

$$(13)\qquad \sigma(\hat{g} - g) = 1/(\sqrt{n}\alpha').$$

5. An example. Using same example outlined in previous paper (see [1, pp. 307-309]), we have $n = 57$, $\hat{\alpha}' = .01924$, $1 - C = .422784$, $y_0 = 4.60015$. This gives $\sigma = 27.826$. For 95% confidence interval we take $(1.96)\sigma = 54.54$, and with $\hat{u} = 180.6$,

$$\hat{g} = \hat{u} + y_0/\hat{\alpha} = 419.7,$$

and the interval is approximated by

$$|\hat{g} - g| < 54.5,$$

which as an approximation gives the symmetrical interval

$$365.2 < g < 474.2.$$

Method 4 used in previous paper gave the longer interval (see Introduction) which was not symmetrical about $\hat{g}$;

$$362.8 < g < 507.4.$$

REFERENCES

[1] B. F. Kimball, "Sufficient statistical estimation functions for the parameters of the distribution of maximum values," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 299-309.

[2] S. S. Wilks, *Mathematical Statistics*, Princeton Univ. Press, 1943, p. 139.

NOTES

This section is devoted to brief research and expository articles and other short items.

TESTS OF INDEPENDENCE IN CONTINGENCY TABLES AS UNCONDITIONAL TESTS

By A. M. Mood

Iowa State College[1]

Summary and introduction. Since the ordinary tests for independence in contingency tables use test criteria whose distributions depend on unknown parameters, the justification for the tests is usually made either by an appeal to asymptotic theory or by interpreting the tests as conditional tests. The latter approach employs the conditional distribution of the cell frequencies given the marginal totals, and was first described by Fisher [1]. The purpose of the present note is to show how these tests may be regarded as unconditional tests even though the parameters are unknown by augmenting the test criterion to include estimates of the unknown parameters. We present no new tests, merely a new setting for the old tests which seems to put them in a little better light.

1. Certain conditional tests. A variate or set of variates x has a probability density function $f(x; \theta)$ under a null hypothesis involving a parameter or set of parameters θ. When the parameters have a set of sufficient estimators $\hat{\theta}$, the joint density function of a random sample of size n may be put in the form

$$\prod_{i=1}^{n} f(x_i ; \theta) = g(x_1 , x_2 , \cdots , x_n \mid \hat{\theta}) h(\hat{\theta}; \theta). \tag{1}$$

It is assumed that n exceeds the number of parameters. We shall be concerned with the class of test criteria which are not functions of the estimators alone. Let $\lambda(x_1 , x_2 , \cdots , x_n)$ be a test criterion which may not be put in the form $\lambda(\hat{\theta})$. The joint density function for λ and $\hat{\theta}$, obtained by summing (1) for fixed λ and $\hat{\theta}$, will be of the form

$$k(\lambda \mid \hat{\theta}) h(\hat{\theta}; \theta). \tag{2}$$

The marginal distribution of λ will be denoted by $m(\lambda; \theta)$, the result of summing (2) over $\hat{\theta}$ for fixed λ.

In order to test the hypothesis in question one would like to divide the λ space into two regions, an acceptance region S_a and a critical region S_c in such a way that S_c would have a prescribed size α under the null hypothesis. One would of course set up other specifications to be fulfilled by S_c, but we are

[1] The author is now with *The RAND Corporation*, Santa Monica, California.

interested here only in the fact that the size of S_c cannot be determined because of the presence of the unknown parameters θ in $m(\lambda; \theta)$.

One can set up a conditional test by using the conditional distribution $k(\lambda \mid \hat{\theta})$. That is, for fixed $\hat{\theta}$, the measure of any region $R(\hat{\theta})$ (which is measurable relative to $k(\lambda \mid \hat{\theta})$, say, in the Lebesgue-Stieltjes sense) of the λ space is known because the $\hat{\theta}$ are known in any given instance. Thus a conditional test can be made with a critical region $R_c(\hat{\theta})$ of prescribed size.

The conditional test may be interpreted as an unconditional test in the present instance in the following manner: the unconditional test is made by using the double criterion $(\lambda, \hat{\theta})$. The $(\lambda, \hat{\theta})$ space is divided into two regions, T_a for acceptance and T_c for rejection. The critical region T_c consists of all points $(\lambda, \hat{\theta})$ such that λ is contained in $R_c(\hat{\theta})$. If the size of $R_c(\hat{\theta})$ is α for all $\hat{\theta}$, then the size of T_c is also α, for

$$\begin{aligned}\int_{T_c} \int k(\lambda \mid \hat{\theta}) h(\hat{\theta}; \theta)\, d\lambda\, d\hat{\theta} &= \int_{-\infty}^{\infty} \left[\int_{R_c(\hat{\theta})} k(\lambda \mid \hat{\theta})\, d\lambda \right] h(\hat{\theta}; \theta)\, d\hat{\theta} \\ &= \int_{-\infty}^{\infty} \alpha h(\hat{\theta}; \theta)\, d\hat{\theta} \qquad (3) \\ &= \alpha.\end{aligned}$$

In this way one can make an unconditional test of the hypothesis with a critical region of prescribed size; of course one does not have complete freedom to specify the shape of T_c, but he can control it to the extent that $R_c(\hat{\theta})$ may be chosen arbitrarily for every $\hat{\theta}$. T_c is of course a similar region in the sense of Neyman and Pearson [2, 3, 4] for the augmented criterion, and the construction of T_c is essentially the same as that used by Neyman and Pearson to test parameters with sufficient estimators.

2. Application to contingency tables. As an illustration we shall follow Wilks' [5] treatment of a two-way table with r rows and c columns; the cell frequencies are n_{ij} and the cell probabilities are p_{ij} with

$$\sum n_{ij} = n; \quad \sum p_{ij} = 1; \quad i = 1, 2, \cdots, r, \quad j = 1, 2, \cdots, c. \qquad (4)$$

The sample is thus regarded as having come from a multinomial population. We let

$$(5) \qquad p_{i\cdot} = \sum_j p_{ij}; \qquad p_{\cdot j} = \sum_i p_{ij}; \qquad n_{i\cdot} = \sum_j n_{ij}; \qquad n_{\cdot j} = \sum_i n_{ij}.$$

The null hypothesis H_0 (of independence) corresponds to the subspace for which

$$(6) \qquad p_{ij} = p_i q_j; \qquad \Sigma p_i = 1 = \Sigma q_j$$

in the parameter space of the p_{ij}. The likelihood ratio criterion for testing H_0 is

$$(7) \qquad \lambda = \frac{(\Pi n_{i\cdot}^{n_{i\cdot}})(\Pi n_{\cdot j}^{n_{\cdot j}})}{n^n \Pi n_{ij}^{n_{ij}}}$$

and its distribution depends on the unknown parameters p_i and q_j. However the parameters have sufficient estimators

$$\hat{p}_i = n_i/n, \qquad \hat{q}_j = n_j/n \tag{8}$$

for the marginal distribution of the $n_{i.}$ and $n_{.j}$ is

$$\frac{(n!)^2}{(\Pi n_{i.}!)(\Pi n_{.j}!)} (\Pi p_i^{n_{i.}})(\Pi q_j^{n_{.j}}) \tag{9}$$

and when this is divided into the distribution of the n_{ij} (under the null hypothesis) one finds the conditional distribution of the n_{ij} to be

$$g(n_{11}, n_{12}, \cdots, n_{rc} \mid n_{1.}, n_{2.}, \cdots, n_{.c}) = \frac{(\Pi n_{i.}!)(\Pi n_{.j}!)}{n! \Pi n_{ij}!} \tag{10}$$

which is independent of the parameters. The distribution (10) is just the combinatorial distribution used ordinarily in deriving the distribution of λ for small samples. The test for independence is therefore a conditional test which however may be interpreted as an unconditional test if the criterion λ is augmented by the estimators of the parameters under the null hypothesis. Instead of the likelihood ratio criterion Karl Pearson's Chi-square criterion could just as well have been used since its conditional distribution is also determined by (10).

The usual difficulty due to discreteness arises in this application to contingency tables. It is not possible to make the significance level exactly α. In terms of the notation of the first section, $R_c(\hat{\theta})$ cannot be chosen so that it will have size exactly equal to α for all $\hat{\theta}$. One would ordinarily replace the equalities by inequalities. The $R_c(\hat{\theta})$ would be chosen to have size less than but as close to α as possible. The size of T_c is then unspecified and one can only state that his significance level is less than α. This difficulty is not particularly serious in practice unless the test criterion has only one degree of freedom.

REFERENCES

[1] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, London, 1946, pp. 96, 97

[2] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. Phil. Trans.*, Series A, Vol. 231 (1933), p. 289.

[3] J. Neyman and E. S. Pearson, "Sufficient statistics and uniformly most powerful tests of statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), p. 113.

[4] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Roy. Soc. Phil. Trans.*, Series A, Vol. 236 (1937), p. 364.

[5] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1943, pp. 213-220.

THE 5% SIGNIFICANCE LEVELS FOR SUMS OF SQUARES OF RANK DIFFERENCES AND A CORRECTION

By Edwin G. Olds

Carnegie Institute of Technology

About ten years ago this author published a paper [1], containing tables for use in testing the significance of the rank correlation coefficient. In a paper on non-parametric tests, [2, p. 316] Scheffé remarks that it would be desirable to have these tables extended by inclusion of the 5% values. When the computation was begun it was noted that a necessary formula was given incorrectly. The main purpose of this note is to correct the formula and to extend Table V, [1, p. 148]. Incidentally, a minor addition for Table III, [1, p. 143] will be supplied.

The formula for the rank correlation coefficient, r', is given by

$$r' = 1 - \frac{6 \sum d^2}{n^3 - n},$$

where n is the number of individuals ranked and $\Sigma d^2 = \sum_{i=1}^{n} d_i^2$ (d_i being the rank difference for the ith individual). As noted in the original paper, the null hypothesis, $r' = 0$, is equivalent to the hypothesis $\Sigma d^2 = (n^3 - n)/6$, and the latter hypothesis is slightly more convenient to test. Scheffé's remark seems to be directed at Table V, which gives, for $11 \leq n \leq 30$, pairs of values between which Σd^2 has a probability, P, of being included. Values are tabled for $P = .99$, .98, .96, .90 and .80. The necessary values for $P = 95$ are given below and can easily be copied in the left-hand margin of the original Table [1, p. 148]. These values, as in the previous case, have been calculated by using the fact that

$$x = \frac{\sum d^2}{2} - \frac{n^3 - n}{12}$$

has an approximately normal distribution with a mean of zero and a variance of $(n - 1)[n(n + 1)/12]^2$. In the original paper, [1, p. 142] the denominator in the bracketed part of the variance was printed as 6, instead of 12.

In this author's original paper the exact frequencies of sums of squares of rank differences were given for $n = 2$ to $n = 7$ inclusive, [1, p. 139]. The same results, together with the results for $n = 8$, were obtained (independently) by Kendall and others and published some months later, [3, p. 255]. Therefore, it is possible to extend slightly the comparison of approximating functions given in Table III, [1, p. 143]. Using Kendall s results for $n = 8$ it is found that when the approximations obtained by using a Pearson Type II curve are compared with exact results the average and maximum differences of cumulatives are .0013 and .0067 respectively. When approximations are made by using the normal curve the corresponding errors are .0081 and .0163.

REFERENCES

[1] E G Olds, "Distribution of the sums of squares of rank differences for small numbers of individuals," *Annals of Math. Stat.*, Vol 9 (1938), pp 133-148.

[2] H Scheffé, "Statistical inference in the non-parametric case," *Annals of Math Stat.*, Vol 14 (1943), pp. 305-332.

[3] M. G Kendall, Sheila F. H. Kendall and B Babington Smith, "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika*, Vol. 30 (1939), pp 251-273

TABLE V (*Extended*)

Pairs of values between which Σd^2 has a probability, P, of being included

n	$P = .95$	
11	83.6	356.4
12	117.0	455.0
13	158.0	570.0
14	207.7	702.3
15	266.7	853.3
16	335.9	1024.1
17	416.2	1215.8
18	508.4	1429.6
19	613.3	1666.7
20	732.0	1928.0
21	865.1	2214.9
22	1013.5	2528.5
23	1178.2	2869.8
24	1360.0	3240.0
25	1559.8	3640.2
26	1778.4	4071.6
27	2016.7	4535.3
28	2275.7	5032.3
29	2556.2	5563.8
30	2859.0	6131.0

INDEPENDENCE OF NON-NEGATIVE QUADRATIC FORMS IN NORMALLY CORRELATED VARIABLES

By Bertil Matérn

Forest Research Institute, Experimentalfältet, Sweden

In a recent paper by the author [5] the following theorem has been mentioned without proof. Though the theorem is very simple and easy to prove the author has not found it elsewhere in the literature.

Theorem. *If two non-negative quadratic forms in normally correlated variables with zero means are uncorrelated the two forms are independent.*

To prove the theorem, let the two forms be

$$(1) \qquad Q_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j, \qquad Q_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} x_i x_j,$$

where the x_i's are normally correlated and all have mean 0. By a well-known theorem on quadratic forms we can reduce Q_1 and Q_2 to the forms

$$(2) \qquad Q_1 = \sum_{i=1}^{n} c_i y_i^2, \qquad Q_2 = \sum_{i=1}^{n} d_i z_i^2,$$

where the y_i's and z_i's are linear functions of the x_i's. In the $2n$-dimensional normal distribution of the y_i's and the z_i's, let ρ_{ij} be the covariance of y_i and z_j. It is then easily shown that the covariance of y_i^2 and z_j^2 is $2\rho_{ij}^2$, and hence that

$$(3) \qquad \operatorname{cov}(Q_1, Q_2) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} c_i d_j \rho_{ij}^2.$$

As the forms are supposed to be non-negative all coefficients in (2) are non-negative. If Q_1 and Q_2 are uncorrelated, each term on the right hand of (3) must vanish. Consequently, if $c_i \neq 0$ and $d_j \neq 0$, we must have $\rho_{ij} = 0$. This means that all y_i's in Q_1 with non-zero coefficients are independent of all z_i's in Q_2 with non-zero coefficients. Hence Q_1 and Q_2 are independent. Q.E.D.

To see if Q_1 and Q_2 are uncorrelated we need an expression for the covariance of the two forms in terms of the coefficients in (1) and the variances and covariances of the original variables x_i. Let A and B be the matrices of the two forms (1). Clearly we may suppose A and B to be symmetric. Let the variance-covariance matrix of the x_i's be L. By straightforward calculations we find

$$(4) \qquad \operatorname{cov}(Q_1, Q_2) = 2\,Tr\,ALBL.$$

Here we have used $Tr\,M$ to denote the "trace," i.e. the sum of the diagonal elements in a square matrix M. In case of independent variables with variance 1, we get

$$(5) \qquad \operatorname{cov}(Q_1, Q_2) = 2\,Tr\,AB.$$

The formulae (4) and (5) are given in [5].

It is interesting to note the simplification of the independence condition given in [2, 3] which is possible when the forms are assumed to be non-negative. It may also be of interest to note that the condition for independence given in the present theorem is identical with the corresponding condition for two linear forms. (In fact, the latter condition has been used in the above proof.) Further we observe that if Q_2 is the square of a linear form with mean 0, we get a necessary and sufficient condition for independence between a linear form and a non-negative quadratic form. The corresponding condition when Q_1 is not supposed to be non-negative has been given in [4].

As an application consider a quadratic form Q in normally correlated variables. Let it be known that Q has a χ^2-distribution with f degrees of freedom. If further

$$Q = Q_1 + Q_2 + \cdots + Q_s, \tag{6}$$

where the Q_i's are non-negative and mutually uncorrelated quadratic forms, then each Q_i has a χ^2-distribution with f_i degrees of freedom, say, and $\Sigma f_i = f$. The proof with the aid of the above theorem is almost immediate. We thus get another formulation of the theorem of Cochran [1] on the decomposition of a quadratic form

REFERENCES

[1] W. G. Cochran, "Distribution of quadratic forms in a normal system with applications to the analysis of covariance," *Proc. Cambr. Phil. Soc.*, Vol. 30 (1934), pp. 178–191.

[2] A. T. Craig, "Note on the independence of certain quadratic forms," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 195–197.

[3] H. Hotelling, "Note on a matric theorem of A. T. Craig," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 427–429.

[4] M. Kac, "A remark on independence of linear and quadratic forms involving independent gaussian variables," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 400–401.

[5] B. Matérn, "Metoder att uppskatta noggrannheten vid linje- och provytetaxering" ("Methods of estimating the accuracy of line and sample plot surveys"), *Meddelanden från Statens Skogsforskningsinstitut*, Vol. 36 (1947), pp. 1–138.

A FORMULA FOR THE PARTIAL SUMS OF SOME HYPERGEOMETRIC SERIES

By Hermann von Schelling

Naval Medical Research Laboratory, New London, Conn.[1]

Let an urn contain N balls of which are a black and b white. A single ball is drawn. We note its color, return the ball into the urn and add Δ balls of the same color. The probability $w(n_1)$ to obtain n_1 black balls in n trials is given by a formula due to F. Eggenberger and G. Pólya [1]:

[1] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Navy Department.

$$(1)\qquad w(n_1)\quad \binom{n}{n_1}\frac{a(a+\Delta)\cdots[a+(n_1-1)\Delta]\cdot b(b+\Delta)\cdots[b+(n-n_1-1)\Delta]}{N(N+\Delta)\cdots[N+(n-1)\Delta]}$$

(n fixed, n_1 variable).

Now, we fix n_1 and ask for the probability that the n_1th black ball appears at the nth drawing. We find

$$(2)\qquad w(n)\quad \binom{n-1}{n_1-1}\frac{a(a+\Delta)\cdots[a+(n_1-1)\Delta]\cdot b(b+\Delta)\cdots[b+(n-n_1-1)\Delta]}{N(N+\Delta)\cdots[N+(n-1)\Delta]}$$

(n_1 fixed, n variable)

This function is the $(n-n_1+1)$th element of the series

$$\frac{a}{\Delta}\left(\frac{a}{\Delta}+1\right)\cdots\left[\frac{a}{\Delta}+(n_1-1)\right]\cdot\frac{\Delta}{N}\left(\frac{N}{\Delta}+1\right)\cdots\left[\frac{N}{\Delta}+(n_1-1)\right]\cdot F\left(n_1,\frac{b}{\Delta},\frac{N}{\Delta}+n_1;1\right).$$

Consequently, the probability that the n_1th black ball appears at the latest in the nth drawing reads, with an obvious abbreviation,

$$(3)\qquad W(n)=\sum_{i=n_1}^{n}w(i)$$
$$\frac{a}{\Delta}\left(\frac{a}{\Delta}+1\right)\cdots\left[\frac{a}{\Delta}+(n_1-1)\right]\cdot\frac{\Delta}{N}\left(\frac{N}{\Delta}+1\right)\cdots\left[\frac{N}{\Delta}+(n_1-1)\right]\cdot F_{n-n_1+1}\left(n_1,\frac{b}{\Delta},\frac{N}{\Delta}+n_1;1\right).$$

Now, we assume the n_1th black ball did not appear in the nth drawing. What is the alternative? The $(n-n_1+1)$th white ball must have appeared in the nth drawing at latest. The corresponding probability is according to the equation (3)

$$(4)\qquad \overline{W}(n)=\sum_{i=n-n_1+1}^{n}\overline{w}(i)$$
$$\frac{b}{\Delta}\left(\frac{b}{\Delta}+1\right)\cdots\left[\frac{b}{\Delta}+(n-n_1)\right]\cdot\frac{\Delta}{N}\left(\frac{N}{\Delta}+1\right)\cdots\left[\frac{N}{\Delta}+(n-n_1)\right]F_{n_1}\left(n-n_1+1,\frac{a}{\Delta},\frac{N}{\Delta}+n-n_1+1;1\right).$$

The relation (4) originates from (3) by writing b instead of a and $(n-n_1+1)$ instead of n_1. The alternatives add to certainty:

$$(5)\qquad W(n)+\overline{W}(n)=1.$$

Change the notations in the following manner:

$$n_1 \to \alpha, \qquad \frac{b}{\Delta} \to \beta; \qquad \frac{N}{\Delta} + n_1 \to \gamma, \qquad n - n_1 + 1 \to \nu. \tag{6}$$

From (6.1) and (6.4) find by addition

$$n \to \nu + \alpha - 1. \tag{7}$$

From (6.1) and (6.3)

$$\frac{N}{\Delta} \to \gamma - \alpha. \tag{8}$$

From (6 2) and (8)

$$\frac{a}{\Delta} = \frac{N}{\Delta} - \frac{b}{\Delta} \to \gamma - \alpha - \beta. \tag{9}$$

Formula (5) reads now

$$\frac{(\gamma-\alpha-\beta)(\gamma-\alpha-\beta+1)\cdots(\gamma-\beta-1)}{(\gamma-\alpha)(\gamma-\alpha+1)\cdots(\gamma-1)} \cdot F_\nu(\alpha,\beta,\gamma;1) + \frac{\beta(\beta+1)\cdots(\beta+\nu-1)}{(\gamma-\alpha)(\gamma-\alpha+1)\cdots(\gamma-\alpha+\nu-1)} \cdot F_\alpha(\nu,\gamma-\beta-\alpha,\gamma-\alpha+\nu;1) = 1. \tag{10}$$

$F_\nu(\alpha,\beta,\gamma;1)$ denotes the partial sum of the first ν elements of the hypergeometric series $F(\alpha,\beta,\gamma;1)$. It is to be mentioned that α is a positive integer necessarily as follows from (6.1). Since

$$W(\infty) = 1 = \frac{(\gamma-\alpha-\beta)(\gamma-\alpha-\beta+1)\cdots(\gamma-\beta-1)}{(\gamma-\alpha)(\gamma-\alpha+1)\cdots(\gamma-1)} F_\infty(\alpha,\beta,\gamma;1),$$

the relation (10) can be written

$$\frac{F_\nu(\alpha,\beta,\gamma;1)}{F_\infty(\alpha,\beta,\gamma;1)} + \frac{F_\alpha(\nu,\gamma-\beta-\alpha,\gamma-\alpha+\nu;1)}{F_\infty(\nu,\gamma-\beta-\alpha,\gamma-\alpha+\nu;1)} = 1, \tag{11}$$

where ν and α are positive integers.

This result is not interesting from the standpoint of pure mathematics since the sum $F(\alpha,\beta,\gamma;1)$ is known. But the relation is useful for the statisticians. In calculating the function $W(n)$ they need a sum of n_1 elements instead of $(n - n_1 + 1)$. If n_1 is small (and this holds in practical applications), the exact calculation of $W(n)$ is possible for every n.

REFERENCES

[1] F. Eggenberger and G. Pólya, *Zeits. f. angew. Math. und Mech*, Vol. 3 (1923), pp. 279–289.

[2] H. von Schelling, *Naturwiss.*, Vol. 30 (1942), pp. 757–758.

THE VARIANCE OF THE PROPORTIONS OF SAMPLES FALLING WITHIN A FIXED INTERVAL FOR A NORMAL POPULATION

By G. A. Baker

University of California, Davis

Suppose that we have a normal population

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \tag{1}$$

and we draw samples of N from this population. We wish to estimate the proportion, p, of the population between two fixed limits, $m + \lambda\sigma$ and $m + \mu\sigma$. One way to make this estimate is simply to count the number of observed x's which fall in this interval. We shall denote this number by n. Then the ratio

$$n/N \tag{2}$$

is an estimate of p. If this is done the variance of p is well known to be

$$\frac{p(1-p)}{N}. \tag{3}$$

The method of estimating p by counting the number in a definite interval is nonparametric and requires no assumption of normal or other specified type of sampled population for validity. However, if we know that the sampled population is normal then we may make use of this knowledge in estimating p and possibly obtain an improved estimate.

Another way to estimate p which makes use of the form of the sampled population is to compute

$$\begin{aligned} \bar{x} &= \frac{1}{N}\sum_{i=1}^{N} x_i \\ s^2 &= \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2, \end{aligned} \tag{4}$$

and hence the integral

$$\int_{m+\lambda s}^{m+\mu s} \frac{e^{-(x-\bar{x})^2/(2s^2)}}{s\sqrt{2\pi}}\, dx. \tag{5}$$

It is implied in elementary texts that (5) is a better estimate of p than is (2) although this point is not discussed.

It is the purpose of the present note to discuss the variance of the estimate (5) and compare this variance with (3).

Now (5) is a function of the first two moments of the sample and it follows from an application of a theorem stated by H. Cramér [1] that (5) is asymptotically normal with mean p and variance given by

$$\sigma_p^2 = \frac{1}{2\pi N}\left[\frac{(\lambda e^{-\frac{1}{2}\lambda^2} - \mu e^{-\frac{1}{2}\mu^2})^2}{2} + (e^{-\frac{1}{2}\lambda^2} - e^{-\frac{1}{2}\mu^2})^2\right]. \tag{6}$$

To compare the relative efficiency of the counting method with (6) in complete detail would be somewhat tedious. The referee suggests a brief discussion of the cases $\lambda = -\infty$, where we are counting the proportion less than some known value, and $\lambda = -\mu$, where a portion out of the middle of the distribution is being counted. These cases are of particular practical interest.

If $\lambda = -\infty$, then (6) becomes

$$\sigma_p^2 = \frac{e^{-\mu^2}}{2\pi N}\left[\frac{\mu^2}{2} + 1\right]. \tag{7}$$

We choose values of μ as indicated below:

μ	p	Relative Efficiency of (3)
−2.3263	0.01	0.27
−1.2816	0.1	0.56
−0.8416	0.2	0.66
−0.5244	0.3	0.75
−0.2533	0.4	0.64
0.0000	0.5	0.64

We get values of the relative efficiency of (3) that are low for small p and somewhat higher for larger values of p.

If $\lambda = -\mu$, then (6) becomes

$$\sigma_p^2 = \frac{\mu^2 e^{-\mu^2}}{\pi N}. \tag{8}$$

We choose values of μ as indicated below:

μ	p	Relative Efficiency of (3)
1.2816	0.8	0.63
0.8416	0.6	0.46
0.2533	0.2	0.12

We see that the relative efficiency of (3) ranges from close to 0.75 to rather small values.

Other choices of λ and μ yield relative efficiencies of about the same order of magnitude as those illustrated.

REFERENCE

[1] Harald Cramér, *"Mathematical Methods of Statistics,"* Princeton University Press, 1946, section 28.4, pp. 366–367

THE POINT BISERIAL COEFFICIENT OF CORRELATION

By Joseph Lev

New York State Department of Civil Service

The product moment coefficient of correlation between a continuous variate y and a variate x which takes the values 1 and 0 only, is known in psychological statistics as the point biserial coefficient of correlation. Let y_i, $i = 1, \cdots, n$, be observations on y; y_{1i}, $i = 1, \cdots, n_1$, be y values which are paired with the value $x = 1$; y_{0i}, $i = 1, \cdots, n_0$, be values paired with $x = 0$; $\bar{y}$, $\bar{y}_1$, and $\bar{y}_0$ be the corresponding means; and $n = n_1 + n_0$. Then the point biserial coefficient of correlation may be written

$$r = \frac{\sqrt{\frac{n_1 n_0}{n}}\,(\bar{y}_1 - \bar{y}_0)}{\left[\sum_{i=0}^{1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2\right]^{\frac{1}{2}}}. \tag{1}$$

The distribution of r is readily obtained when the y_i, $i = 1, \cdots, n$, are distributed as

$$\frac{1}{\sqrt{2\pi}\,\sigma\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2\sigma^2(1-\rho^2)}(y_i - \alpha - \rho\sigma z_i)^2\right] \tag{2}$$

where

$$z_i = \frac{x_i - \bar{x}}{\sigma_x} = \begin{cases} \sqrt{\dfrac{n_0}{n_1}}, & i = 1, 2, \cdots, n_1, \\ -\sqrt{\dfrac{n_1}{n_0}}, & i = n_1 + 1, n_1 + 2, \cdots, n, \end{cases}$$

σ^2 is the variance of the y_i about the common mean α, and ρ is the parameter which represents the correlation between the y_i and the x_i. It is easy to verify that the statistic in (1) is a maximum likelihood estimate of ρ.

It will be convenient to express the two population means in (2) as μ_1 and μ_0 so that

$$\begin{aligned} \mu_1 &= \alpha + \rho\sigma\sqrt{\frac{n_0}{n_1}}, \\ \mu_0 &= \alpha - \rho\sigma\sqrt{\frac{n_1}{n_0}}. \end{aligned} \tag{3}$$

Hence

$$\rho = \sqrt{\frac{n_1 n_0}{n}}\,\frac{\mu_1 - \mu_0}{\sigma}. \tag{4}$$

Now write

$$t = \frac{\sqrt{\frac{n_1 n_0}{n}}(\bar{y}_1 - \bar{y}_0)\sqrt{n-2}}{\left[\sum_{i=0}^{1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2\right]^{\frac{1}{2}}} = \frac{\sqrt{n-2}\,r}{\sqrt{1-r^2}}, \tag{5}$$

where r is obtained from (1).

Using (5) we may write t as

$$t = \frac{\dfrac{(\bar{y}_1 - \bar{y}_0) - (\mu_1 - \mu_0)}{\sqrt{\frac{n}{n_1 n_0}}\sigma\sqrt{1-\rho^2}} + \dfrac{\mu_1 - \mu_0}{\sqrt{\frac{n}{n_1 n_0}}\sigma\sqrt{1-\rho^2}}}{\dfrac{\left[\dfrac{\sum_{i=0}^{1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n-2}\right]^{\frac{1}{2}}}{\sigma\sqrt{1-\rho^2}}}.$$

Therefore t has non-central t distribution [1] with

$$\delta = \frac{\mu_1 - \mu_0}{\sqrt{\frac{n}{n_1 n_0}}\sigma\sqrt{1-\rho^2}} = \sqrt{n}\,\frac{\rho}{\sqrt{1-\rho^2}}. \tag{6}$$

The methods and tables given in [1] may be used to calculate tests of significance and confidence limits for ρ.

When $\rho = 0$, t has Student's distribution, and the statistic $t = \sqrt{n-2}\,r/\sqrt{1-r^2}$ may be used to test the hypothesis, $\rho = 0$, by means of the t tables with $n - 2$ degrees of freedom. The non-central t distribution then determines the power function of this test.

Table IV of [1] can be used to calculate confidence limits for ρ. If the confidence interval is to be based on equal tails of the distribution choose a confidence coefficient $1 - 2\epsilon$. Then compute $\delta(f, t_0, \epsilon)$ and $\delta(f, t_0, 1-\epsilon)$, where $f = n - 2$, and $t_0 = \sqrt{n-2}\,r/\sqrt{1-r^2}$.

A lower limit for ρ is given by

$$\frac{\delta(f, t_0, \epsilon)}{[n + \delta^2(f, t_0, \epsilon)]^{\frac{1}{2}}},$$

and an upper limit by

$$\frac{\delta(f, t_0, 1-\epsilon)}{[n + \delta^2(f, t_0, 1-\epsilon)]^{\frac{1}{2}}}.$$

REFERENCE

[1] N. L. Johnson and B L Welch, "Applications of non-central t-distribution," *Biometrika*, Vol. 31 (1940), pp 362–389.

A NOTE ON KAC'S DERIVATION OF THE DISTRIBUTION OF THE MEAN DEVIATION

By H. J. Godwin

University College of Swansea, Wales

In a paper on a general class of estimates of deviations, Kac [3] obtained an expression for the frequency function of the estimate of mean deviation from the mean in normal samples. He was unable to establish the identity of this with an expression obtained earlier by me [1]. I now shew that the two results are, in fact, equivalent.

Kac uses the functions $\zeta^{(k)}(x)$, defined as the k − fold convolution of

$$\zeta(x) = \begin{cases} 0, & x < 0; \\ e^{-\frac{1}{2}n^2x^2}, & x \geq 0. \end{cases}$$

I used the functions $G_k(x)$ defined by the recurrence relation

$$(1) \qquad G_0(x) = 1, \qquad G_k(x) = \int_0^x e^{-(t^2/2k(k+1))}\, G_{k-1}(t)\, dt$$

Now I have shewn elsewhere [2] that the integral of $e^{-\frac{1}{2}(x_1^2+\cdots+x_k^2)}$ taken through the interior of a regular simplex in k dimensions, with its centroid at the origin and of side a, is $\sqrt{k+1}\, G_k(a/\sqrt{2})$. The relation (1) corresponds to a dissection of the simplex into sections, which are $(k-1)$-dimensional simplexes, by joining the centroid to the vertices and taking sections parallel to the base of each of the $(k+1)$ smaller simplexes so formed. If however we take sections parallel to a base of the whole simplex we get another recurrence relation, viz.

$$(2) \qquad G_k(x) = \int_0^x e^{-((kx-(k+1)t)^2/2k(k+1))} G_{k-1}(t)\, dt\,.$$

Now (2) may be re-written

$$\frac{G_k(nx)e^{-(n^2x^2/2(k+1))}}{n^k} = \int_0^x e^{-(n^2(x-t)^2/2)}\, \frac{G_{k-1}(nt)e^{-(n^2t^2/2k)}}{n^{k-1}}\, dt$$

whence, by induction, $G_{k-1}(nx)\cdot e^{-(n^2x^2/2k)} = n^{k-1}\zeta^{(k)}(x)$ and the equivalence of Kac's result to mine is established.

REFERENCES

[1] H. J. Godwin, "On the distribution of the estimate of mean deviation obtained from samples from a normal population," *Biometrika*, Vol. 33 (1945), pp. 254–256.

[2] H. J. Godwin, "A further note on the mean deviation," *Biometrika*, Vol. 35 (in the press).

[3] M. Kac, "On the characteristic functions of the distributions of estimates of various deviations in samples from a normal population," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 257–261.

CORRECTION TO "ASYMPTOTIC FORMULAS FOR SIGNIFICANCE LEVELS OF CERTAIN DISTRIBUTIONS"

By A. M. Peiser

New York City

Professor Henry Scheffé has recently pointed out to me an error in my paper "Asymptotic formulas for significance levels of certain distributions," which appeared in *Annals of Math. Stat.*, Vol. 14 (1943), pp. 56–62. In the determination of the significance levels of Student's t distribution, appeal was made to a theorem of Cramér which requires independent random variables. The variables defined at the top of page 61, however, *cannot* be taken as independent, so that the theorem does not apply.

The asymptotic formula (following the notation of the paper)

$$t_{p,n} = y_p + \frac{y_p^3 + y_p}{4n} + o\left(\frac{1}{n}\right),$$

where

$$\Phi(y_p) = 1 - p, \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-v^2/2}\, dv,$$

is nevertheless correct. This may be shown directly from the distribution function

$$G_n(x) = \tfrac{1}{2} + \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \int_0^x \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt.$$

Writing

$$\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} = \exp\left[-\frac{n+1}{2} \log\left(1 + \frac{t^2}{n}\right)\right]$$

$$= \exp\left[-\frac{n+1}{2}\left(\frac{t^2}{n} - \frac{t^4}{2n^2} + \frac{t^6}{3(n+\theta t^2)^3}\right)\right], \qquad |\theta| < 1,$$

and using Stirling's formula, it follows that $G_n(x)$ can be written in the form

$$G_n(x) = \tfrac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} \left[1 + \frac{t^4 - 2t^2 - 1}{4n} + \frac{1}{n^2} Q_n(t)\right] dt$$

$$= \Phi(x_p) - \frac{x^3 + x}{4n\sqrt{2\pi}}\, e^{-x^2/2} + \frac{1}{n^2\sqrt{2\pi}} \int_0^x Q_n(t) e^{-t^2/2}\, dt,$$

where $Q_n(t)$ is a bounded function of t and n in each finite interval.

Let $t_{p,n} = y_p + a_n$, where $a_n = o(1)$. Then $G_n(t_{p,n}) = \Phi(y_p) = 1 - p$, and we have

$$na_n \frac{\Phi(y_p + a_n) - \Phi(y_p)}{a_n} = \frac{(y_p + a_n)^2 + (y_p + a_n)}{4\sqrt{2\pi}} e^{-\frac{1}{2}(y_p+a_n)^2} + O\left(\frac{1}{n}\right),$$

so that

$$\lim_{n\to\infty} na_n = \frac{y_p^3 + y_p}{4}.$$

This is the required result.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Seattle Meeting of the Institute on November 26-27, 1948)

1. Estimation of the Variance of the Bivariate Normal Distribution.

HARRY M. HUGHES, University of California, Berkeley.

Let x_1 and x_2 be two random variables normally distributed with known means m_1 and m_2, and with common unknown variance σ^2. Consider an experiment in which the observed variable is $Y = \sqrt{(x_1 - m_1)^2 + (x_2 - m_2)^2}$. This paper considers the problem of estimating the parameter σ when the observations are grouped. By the method of minimum reduced chi-square with linear restrictions, two best asymptotically normal estimates are derived. By minimization of the asymptotic variance of these estimates, the optimum choice of grouping is found as a function of σ. For two and for three groups, when it is known or assumed a priori that σ is on a certain finite interval, the optimum grouping is derived which will minimize the maximum asymptotic variance on that interval. If such interval is moderately small, it is shown that the optimum grouping is the same as if σ were known to have the value at the upper end of the interval. Finally the effect of using non-optimum grouping is analyzed.

2. Derivation of a Broad Class of Consistent Estimates.

R. C. DAVIS, Inyokern, California.

Given a chance vextor X with cumulative distribution function $F(X, \theta)$, where θ is an unknown parameter vector, a broad class of estimates of θ is derived having the following properties: a) any estimate in this class is a consistent estimate of θ; b) any estimate is a symmetric function of independent observations of the chance vector X. The novel feature of this class is that no assumptions about the existence of various partial derivatives of a density function with respect to θ are made. As a matter of fact not even the existence of a density function is required, and it is postulated merely that a continuous function of X for each θ (in a certain neighborhood of the true θ_0) and of θ for each X exist which satisfies a Lipschitz condition in θ. For each such function having a finite first moment an estimate of θ is constructed which has the properties a) and b) listed above.

3. Locally Best Unbiased Estimates.

EDWARD W. BARANKIN, University of California, Berkeley.

Let $p = \{p_\vartheta(x); \vartheta \in \Theta\}$ be a family of probability densities in the space Ω of points x; and g a function on Θ. Let s be fixed and >1; call an unbiased estimate of g *best at* ϑ_0 if its s-th absolute central moment (s.a.c.m.) under p_{ϑ_0} is (finite and) not greater than the s.a.c.m., under p_{ϑ_0}, of any unbiased estimate of g. With a certain integrability postulate on the p_ϑ's, a necessary and sufficient condition, of finite character, is established for the existence of an unbiased estimate of g having a finite s.a.c.m. under p_{ϑ_0}. When such a one exists, there then exists a *unique* unbiased estimate which is best at ϑ_0. The existence condition defines the s.a c.m. of the best estimate explicitly as the l.u.b. of a set of numbers; in particular, we obtain immediately a generalization of the Cramér-Rao inequality. Also, when it exists, the best unbiased estimate is explicitly constructed from the function g and the densities p_ϑ. The case $s = 2$ is studied more closely. Also, a detailed example is considered

4. Some Problems Related to the Distribution of a Random Number of Random Variables. EDWARD PAULSON, University of Washington, Seattle.

Let $\{x_i\}(i = 1, 2, 3 \cdots)$ be a set of independent random variables with identical distributions, with $E(x) = a$ and $\sigma^2(x) = b\ (0 < b < \infty)$ Let N be a positive integral-valued random variable with distribution $F_\lambda(N)$ depending on a parameter λ, where $E(N) = A_\lambda$, and $\sigma^2(N) = B_\lambda\ (0 \leq B_\lambda < \infty)$. Now let $T_N = x_1 + x_2 + \cdots + x_N$ The limiting distribution as $\lambda \to \infty$ of $\dfrac{T_N - aA_\lambda}{\sqrt{a^2B_\lambda + bA_\lambda}}$ has been investigated by Robbins (*Proc. Nat. Acad. Science*, Vol. 34 (April 1948), pp. 162-163) for several different sets of conditions on the distribution of N. It can be shown that analogous results will hold if instead of T_N we consider a more general statistic T_N^*, whose conditional distribution with respect to the variable N is such that there exist constants a_1 and b_1 so that

$$\lim_{N\to\infty} E\left\{\exp\left[it\left(\frac{T_N^* - a_1N}{\sqrt{b_1N}}\right)\right]\right\} = h(t)$$

uniformly in any finite t interval, where $h(t)$ is a characteristic function. Returning to the statistic T_N, it can be shown that there exists an asymptotic expansion in powers of $\lambda^{-\frac{1}{2}}$ with remainder $O(\lambda^{-(k/2)+1})$ for $P\left\{\dfrac{T_N - aA_\lambda}{\sqrt{a^2B_\lambda + bA_\lambda}} \leq u\right\}$ when the following conditions are satisfied. (1) the distribution function of x has a non-zero absolutely continuous component, (2) $E(|x|^k) < \infty$, and (3) $\lambda \to \infty$ through integral values, and $F_\lambda(N)$ is the λth convolution of a random variable n such that $E(n^k) < \infty$

5. Asymptotic Expansions for the Distribution of Certain Likelihood Ratio Statistics. ALBERT H. BOWKER, Stanford University.

Asymptotic expansions of the "Cramerian" type are derived for the distribution of likelihood ratio statistics given by Wilks for testing various hypotheses about means, variances, and covariances of a normal multivariate distribution. The point of departure is Wilks' result that minus twice the logarithm of the likelihood ratio has the χ^2 distribution; terms in $\dfrac{1}{N}, \dfrac{1}{N^2}, \cdots$ may also be expressed in terms of the χ^2 distribution. In addition, asymptotic expansions of the "Fisher-Cornish" type are obtained for the percentage points and for a transformation of the statistic to a χ^2 variate.

6. On a Problem of Confounding in Symmetrical Factorial Design. ESTHER SEIDEN, University of California, Berkeley.

Let $m_3(r, s)$ be the maximum number of factors that is possible to accommodate in symmetrical factorial experiment in which each factor is at $s = p^n$ levels (p being any positive prime number, n a positive integer) and each block is of size s^r, without confounding any degrees of freedom belonging to any interaction involving 3 or lesser number of factors.

R. C. Bose proved in a paper "Mathematical theory of factorial design," *Sankhyā*, Vol. 8 (1947), pp. 107-166, that the following inequality holds.

$s^2 + 1 \leqq m_3(4, s) \leqq s^2 + s + 2$. This gives in case $s = 4$, $17 \leqq m_3(4, 4) \leqq 22$. It is now proved that $m_3(4, 4) = 17$.

The proof consists in showing that the maximum number of non three collinear points which can be chosen in a finite projective space $PG(3, 4)$ cannot exceed 17, which according to a proof of R. C. Bose is equivalent to the staement that $m_3(4, 4)$ cannot exceed 17.

7. Some Bounded Significance Level Tests of Whether the Largest Observations of a Set are Too Small (*Preliminary Report*) JOHN E. WALSH, Santa Monica, California

A set of n observations are given which satisfy (1) They are independent and from continuous symmetrical populations; (2) The r largest observations are from populations with median θ while the remaining observations are from populations with median φ. It is required to test whether $\theta < \varphi$. Let $x(1), \cdots, x(n)$ denote the observations arranged in increasing order of magnitude. For $r = 1$ tests of the form. *Accept* $\theta < \varphi$ *if* $x(n) < 2x(w_\alpha) - x(i)$, *where* $\alpha = \Pr[x(n) + x(i) < 2\theta \mid \theta = \varphi]$ *and* w_α *is the smallest integer satisfying* $\Pr[x(w_\alpha) > \theta \mid \theta = \varphi] \leq \alpha$, can be obtained from $n \geq 15$. Exact significance levels can be obtained by assuming a sample from a specified population (e.g. normal). On the basis of (1)–(2) alone, the significance level never exceeds 2α. For large n, tests can be obtained for any r if the observations satisfy the additional weak condition. (3) The tail order statistics are approximately independent of the central order statistics; also the variances of the tail order statistics are either very large or very small compared with the variances of the central order statistics. The test is: *Accept* $\theta < \varphi$ *if* $\max[x(i_k) + x(n - j_k)]; 1 \leq k \leq s \leq r] < 2x(w_\alpha)$, *where* $i_u < i_{u+1}, j_v < j_{v+1}, j_s = r - 1, w_\alpha$ *is the smallest integer satisfying* $\Pr[x(w_\alpha) > \theta \mid \theta = \varphi] \leq \alpha$, *and* $\alpha = \Pr[\max[x(i_k) + x(n - j_k)] < 2\theta \mid \theta = \varphi]$. For large n the significance level is approximately α but is $\leq 2\alpha$ for all n. The power function $\to 1$ as $\varphi - \theta \to \infty$ and $\to 0$ as $\varphi - \theta \to -\infty$.

8. Determination of Optimal Test Length to Maximize the Multiple Correlation. PAUL HORST, University of Washington, Seattle.

If the lengths of the tests in a battery are altered, their intercorrelations and their validities or correlations with a criterion are also altered. Consequently, the multiple correlation of the battery with the criterion will also be altered. These changes are a function of the reliabilities of the tests. Suppose we have given from a set of experimental data (1) the time allowed for each test in the battery, (2) the reliability of each test, (3) the intercorrelations, and (4) the validities of all the tests. If we specify the overall testing time we are willing to allow for the test in the future, we can determine the amount by which each test must be altered in order to give the maximum multiple correlation with the criterion. The method, together with numerical examples and the mathematical proof, is presented.

9. Some Numerical Comparisons of a Non-Parametric Test with other Tests. F. J. MASSEY, University of Oregon, Eugene.

Let $F(x)$ be the cumulative distribution function of a R.V. X, and let $x_1 < x_2 \cdots < x_n$ be the results of n independent observations ordered as to size.

Define $S_n(x) = 0$ if $x < x_1$;

$$= \frac{k}{n} \text{ if } x_k \leq x < x_k + 1,$$

$$= 1 \text{ if } x_n < x.$$

To test the hypothesis $H_0 : F(x) = F_0(x)$, where $F_0(x)$ is completely specified, use the criterion reject H_0 if $\max \mid S_n(x) - F_0(x) \mid > \frac{\lambda}{\sqrt{n}}$. Choice of λ determines the first kind of error. The second kind of error against specified alternatives can be calculated numerically

10. **On the Deviation of Extreme Values.** W. J. Dixon, University of Oregon, Eugene.

Let $x(i)$ be the ith observation in order of magnitude in a sample of size n. The distribution of $R = \dfrac{x(n) - x(2)}{x(n) - x(1)}$ is obtained explicitly for samples from a rectangular distribution and for $n = 3, 4, 5$, for samples from a normal distribution. Percentage values of R for values of n up to 30 are presented. Generalizations of R are indicated.

11. **The Optimum Size of Interval for Making Measurements of a Rocket's Angular Velocity.** Edward A. Fay, University of California, Berkeley.

Over a given range of time $0 \leq \tau \leq T$, the angular velocity of a rocket's spin is adequately represented by a polynomial $\xi(\tau)$ of given degree $s - 1$ but with unknown coefficients. The rocket's angular acceleration and the angle through which it spins in a given time-interval may then be obtained respectively by differentiating and integrating $\xi(\tau)$. Let ν be an integer $\geq s$, let $t = T/\nu$, and let η_i be the angle through which the rocket turns in the interval $(i - 1)t \leq \tau \leq it$. While $\xi(\tau)$ and $\xi'(\tau)$ cannot be directly observed, the angles $\eta_1, \eta_2, \cdots, \eta_\nu$ can. Let Y_i be an observation on η_i, and assume that $Y_1, Y_2, \cdots, Y_\nu$ are independent homoscedastic variables. The Y_i may then be combined by the method of least squares to obtain best linear estimates $X(\tau, t)$ and $X'(\tau, t)$ of $\xi(\tau)$ and $\xi'(\tau)$. The choice of t is at the observer's disposal. For the cases $s = 2, 3, 4$, and for the cases that the common variance of the Y_i is (a) independent of t or (b) proportional to t, an expression is derived for the variance of $X'(\tau, t)$, and the maximum value of that variance over the range of τ is minimized with respect to t. The method is of much more general application.

12. **Stationary Time Series Analysis and Common Stock Price Forecasting.** Zenon Szatrowski, University of Oregon, Eugene.

The objective of this paper is to present a statistical procedure of practical value in the problem of extracting information from the past behaviour of economic time series, information to be used in projecting future patterns. The author feels that his approach yields results closer to reality than the techniques described by Herman Wold, M. C. Kendall, H. T. Davis, and in particular, the technique of "disturbed harmonics" used by G. U. Yule. The idea of the proposed technique can be described by examining the autoregression scheme, which seems to be considered the most desirable by the above men. A simple example of such a scheme is the equation

$$u_{t+2} = -au_{t+1} - bu_t + E_{t+2},$$

where the u's are the time series values and E's are random elements. The above linear relationship, when determined either directly or through an empirical correlogram (for which data is usually inadequate) is a kind of an average relationship. It may be as inappropriate in estimating future values of a time series as would be an average in estimating the level of a series with a pronounced trend.

The author proposes using derived time series to shed light on the nature of the changes in the parameters under consideration. Such derived series could be estimates of the a's and b's for successive time periods. The author has found that projections of common stock price fluctuations were improved considerably when the changing nature of the cyclical pattern was taken into account. This was done by constructing derived time series, "moving" estimates of the amplitude, period and phase of the dominant harmonics.

The author points out that the above approach has shown promise in commodity prices as well as common stocks The value of this approach in forecasting lies in the facts that (1) it does not require forecasts of other series and (2) it is based on the realistic assumption that history repeats itself but with variations, variations which may be taken into account through appropriate models

13. Distribution of the Number of Schools of Fish Caught Per Boat.

J. NEYMAN, University of California, Berkeley.

Let λ be the average number of schools of fish per unit area of a fishing ground A. Let a be any area partial to A, and let $\Omega(m, a, \lambda)$ denote the probability that exactly n schools of fish will be found within a. At time $t = 0$ a boat begins scouting for fish in A traveling at constant speed v. It is assumed that all schools of fish within distance r of the boat are detected and none is detected at a greater distance. If $s \geqq 1$ schools are detected then they are caught in turn, the catching of one school taking up exactly h hours. $X(t)$ denotes the random variable representing, for each $t \geqq 0$, the number of schools caught up to time t including the one which may be in the process of being caught at the moment t. Probability distribution of $X(t)$ is given by the formula

$$P\{X(t) \leq k\} = \sum_{m=0}^{k} \Omega[m, 2rv(t - kh), \lambda]$$

for $k = 0, 1, 2, \cdots, n - 1$, where $n - 1$ is the greatest integer smaller than t/h. Of course $P\{X(t) \leq n\} = 1$. This result is easily generalized for the joint distribution of catches of several boats fishing in the same area so that their paths do not cross. Assuming specific functions to represent $\Omega(m, a, \lambda)$ formulae may be obtained to estimate the parameters λ and rv.

14. Some Problems in Fishery Research to which Statistical Methods are Applicable (*Preliminary Report*).

RALPH P SILLIMAN, U. S Fish and Wildlife Service, Seattle.

One of the most difficult problems is the obtaining of a random sample of a fish population Rarely are such populations randomly distributed over any area, and the samples must often be taken from the catches of fishing vessels which do not uniformly cover even a part of the area of distribution of the population. Many distributions of variables found in fishery research are not normal, and statistical methods based on the normal distribution can be applied only through the use of unsatisfactory transformations. Since fishery research is largely observational in technique, data reflecting the concurrent effect of several variables are usually obtained Although the present methods of multiple correlation and regression can be used in some instances to measure the relative effect of the separate variables, there are many situations in which these methods do not yield good results. Finally, many data used in fishery research must be adjusted before use, and existing methods do not give good measures of the expected variability of such adjusted data. Examples of specific problems are found in the distribution of deliveries and the variations in catch of Columbia River chinook salmon.

15. The Application of the Hypergeometric Distribution to Problems of Estimating and Comparing Zoological Population Sizes.

DOUGLAS CHAPMAN, University of California, Berkeley.

Estimates and tests of the χ^2 type, as developed by Neyman, are adapted to sampling without replacement from a finite population. These results are applied to problems of

estimation and comparison of zoological population sizes as determined by sampling procedures. For single samples the bias and variance of different estimators is compared. Finally some numerical calculations are made for various population and sample sizes to determine how different sample sizes and different methods of analysis affect the size of the critical region which is necessarily an approximation to the desired size. For some of these the power of the test is considered.

16. Extension to Multivariate Case of Neyman's Smooth Test with Astronomical Application.

ELIZABETH L. SCOTT, University of California, Berkeley.

It is more or less generally accepted that the distribution of extra-galactic nebulae in space is not uniform in the small. In particular, counts in small cubes show distinct signs of contagion. On the other hand, it is not settled whether or not lack of uniformity in the large exists. One way of making this statement precise is to assert that the power series expansion of the logarithm of the probability density of the two angular coordinates of the nebulae within a given large area on the unit sphere does not contain low order terms. In fact, any such low order terms could be interpreted as determining "trends" or what could be described as lack of uniformity in the large. From this point of view, uniformity in the large may be tested by a two dimensional Neyman Smooth Test for goodness of fit.

Let $\{\pi_{ij}(x, y)\}$ be a sequence of polynomials in x and y ortho-normal for $|x| < a$ and $|y| < b$. If x_k and y_k are the coordinates of the kth out of n nebulae counted within the rectangle $(-a, a), (-b, b)$ then the smooth test of mth order consists of rejecting the hypothesis of uniformity in the large when $\sum_{i+j=1}^{m} \left(\sum_{k=1}^{n} \pi_{ij}(x_k, y_k) \right)^2 \geqq n\chi_\epsilon^2$ where χ_ϵ^2 is the tabled value of χ^2 with $m(m+3)/2$ degrees of freedom.

17. A Mathematical Theory of Vitamin A Metabolism in Fish (*Preliminary Report*).

NORMAL E. COOKE, Vancouver, B.C.

Several possible hypotheses for vitamin A metabolism in fish are developed from simple postulates. These hypotheses are tested (by least squares method) against experimental data in an attempt to deduce the correct mechanism.

18. The Interactance Hypothesis between Populations.

STUART C. DODD, University of Washington, Seattle.

The hypothesis of interacting between human populations, or of demographic gravitation, is that the number of interactions between two communities (or other groups) tends to vary directly with the product of the two populations and their "specific coefficients" and the overall duration and tends to vary inversely with the intervening distance and the average duration of an interact. The hypothesis is tested by isolating factors and measuring their correlation with the amount of interacting in the pairs of a set of N communities.

This hypothesis is supported by studies of telephoning; news circulating; travel by bus, train, or plane; R. R. express; college attandance; intermarrying; etc. Further lists of interhuman actions are suggested for investigation.

A new corroborating bit of data comes from a poll by the Washington Public Opinion Laboratory in a Seattle housing project where negro-white relations threatened violence. The tension units of verbal interaction (defined as one anti-negro opinion asserted by one white person) were observed to decrease inversely with a power of the distance from a rape site. The observed tension correlated with the formulas or curves predicting that tension at $\rho = .94$ and passed the chi-square test at the one per cent level. The tension is dimensionally analyzed as a social force and social energy.

19. **The Employment of Marked Members in the Estimation of Animal Populations.** MILNER B. SCHAEFER, U. S Fish and Wildlife Service, Honolulu, T. H

The estimation of population numbers by marked members is an important technique in fisheries research The number N of individuals in the population, of which T are known to be marked, may be estimated from a sample of n of which t are found to be marked. Several estimates are available, all of which reduce to $N = \frac{nT}{t}$ when the numbers are all large, but more precise formulae should be used when the numbers are not all large. An estimate of the variance of N has been derived by Karl Pearson (*Biometrika*, Vol. 20 (1928), pp. 149–174) on the basis of inverse probabilities. The sampling error may also be measured by means of confidence intervals Formulae have been developed for estimating N from repeated samples of the same population, but no very suitable estimates of the sampling error are available in this case. For some migratory fishes marked at a point on their migration path and sampled later at another point, there exists a correlation between time of marking and time of recovery in the subsequent samples. In such case, the total number of fish marked or drawn in the subsequent samples cannot in general be regarded as random samples of the population. Where numbered tags are employed as marks, so the fish may be individually identified both when marked and recovered, a method of estimating N in this case also is suggested

20 **Non-Response and Repeated Call-Backs in Sampling Surveys.** Z. W. BIRNBAUM AND MONROE G. SIRKEN, University of Washington, Seattle.

In opinion-polls and other sampling surveys, a response can only be obtained from those individuals of a sample who are available for interviewing. Let $p_{.1}$ be the probability that an individual chosen at random from the population answers "yes" to a question, $p_{1.}$ that an individual is available for interviewing, and p_{11} that an individual is available and answers "yes." Usually one wishes to estimate the parameter $p_{.1}$, but from a sample it is only possible to estimate $\frac{p_{11}}{p_{1.}} = p' =$ the probability that an individual answers "yes" if he is available. Thus the total error in estimating $p_{.1}$ from a sample contains two components: the bias $p_{.1} - p'$ and the sampling error. In this paper a technique is presented in which individuals not available at a call are called upon repeatedly, up to k times. It is shown how, for a given upper bound of the total error at a prescribed probability level and a given k, it is possible to minimize the cost of the survey by optimizing the relationship between the greatest possible bias and the sampling error

---

(Abstracts of papers presented at the Cleveland Meeting of the Institute on December 27-30, 1948.)

21. **A Necessary Condition for a Certain Class of Characteristic Functions** (*Preliminary report*). EUGENE LUKACS, NOTS, Inyokern, California and Our Lady of Cincinnati College, Cincinnati, Ohio.

Let $\varphi(t) = \left\{\left(1 - \frac{t}{v_1}\right)\left(1 - \frac{t}{v_2}\right)\cdots\left(1 - \frac{t}{v_n}\right)\right\}^{-1}$ be the reciprocal of a polynomial without multiple roots. The following necessary condition is derived which $\varphi(t)$ has to satisfy in order to be the Fourier transform (characteristic function) of a distribution.

If $\varphi(t)$ is the Fourier transform of a distribution, then

1) $\varphi(t)$ has no real roots. If $b + ia$ $(a \neq 0, b \neq 0)$ is a root then $-b + ia$ is also a root. That is the roots of $\varphi(t)$ are either located on the imaginary axis or are symmetrical to this axis.

2) If $b + ia$ $(a \neq 0)$ is a root then there exists also at least one root $i\alpha$ so that sign α = sign a and $|\alpha| \leq |a|$.

As a particular case one obtains the well known fact that $(1 + t^4)^{-1}$ cannot be a characteristic function.

22. Precision of Estimates from Samples Selected under Marginal Restrictions. (*Preliminary Report*). CLIFFORD J. MALONEY, Camp Detrick, Frederick, Maryland.

Formulas are derived for estimates and for their variances computed from samples drawn at random subject only to marginal restrictions from populations classified by several characters, and estimates are made of the efficiency of such sampling plans compared to sampling with complete stratification or sampling completely at random. By means of two simple but general theorems it is shown that the variances are independent of the individual values of the character being sampled for in the population and in the sample and depend only on the first two moments for each cell of the population. It is shown that in the large sample approximation a practical scheme for actually drawing such samples can be obtained by drawing a sample of size n entirely at random and using the results of Deming and Stephan (*Annals of Math. Stat.*, Vol. 11 (1940), p. 427) to adjust the sample marginal totals to the specified values. Deficient cells will of course be filled up by additional drawings. A measure is given of the relative loss of information in sampling with marginal restrictions on the sample cell numbers compared to sampling with complete stratification. If a_{ij} represents the population mean in the ijth cell, r_i the population mean in the ith row and c_j the population mean in the jth column, and if a_{ij} is of the form $a_{ij} = a + r_i + c_j$, then marginally restricted sampling is as efficient as sampling with complete stratification. For arbitrary a_{ij} a measure of the relative efficiency compared to sampling completely at random is given by the relative degrees of freedom for the sample cell numbers. A comparison with other possible sampling procedures is given.

23. Properties of Maximum- and Quasi-Maximum Likelihood Estimates of Parameters of a System of Linear Stochastic Difference Equations with Serially Correlated Disturbances (*Preliminary Report*). HERMAN RUBIN, Cowles Commission, The University of Chicago.

Let $A_{us}x_t' = u_t'$ be a complete system of linear stochastic difference equations, $x_t = (y_t, z_t)$, y_t jointly dependent, z_t predetermined. Let us suppose $u_t' + B_{su}u_{t-1}' = v_t'$, where the random vectors v_t are serially independent and have mean zero. If the vectors v_t have the same Gaussian distribution, and the system is identified, we can obtain maximum-likelihood estimates; if the distributions are not identical Gaussian, quasi-maximum-likelihood estimates result. The identification problem is a special case of that with independent u_t and bilinear restrictions on some A_{us}^*, if the restrictions on A_{us}^* are linear or bilinear. As in that case, we may have multiple identification. However, the special aspects of this type of system yield some help in the discussion of the identification problem. We also observe that if the system is identified, we obtain consistency and asymptotic normality of the estimates under the same conditions as with serially independent u's for A_{us}.

24. The Computation of Maximum Likelihood Estimates of Parameters of a System of Linear Stochastic Difference Equations with Serially Correlated

Disturbances. HERMAN CHERNOFF, Cowles Commission, The University of Chicago

Consider the structural equations $A_{uz}x_t' = u_t'$ where the vector $x_t = (y_t, z_t)$, y_t are the jointly dependent, and z_t the predetermined variables and where u_t are serially correlated. In particular assume that the disturbances u_t satisfy the simple Markoff Process $u_t' + B_{vu}u_{t-1}' = v_t'$ where v_t is a stationary serially uncorrelated Gaussian Process with zero mean. Then we have $A_{uz}x_t' + B_{vu}A_{uz}x_{t-1}' = v_t'$. The estimates of B_{vu} and $E\{v_t'v_t\}$ can be simply expressed in terms of those of A_{uz}. It is shown that iterative gradient methods of maximization require about 2 to 3 times as much work per iteration as in the serially uncorrelated case. To apply the Newton Method about 8 times as much work per iteration is required. The Newton Method uses the second order terms of the expansion of the log of the likelihood in terms of the independent parameters of A_{uz} and these can be used to obtain estimates of the asymptotic covariance of the estimates.

25. Test Criteria for Hypotheses of Symmetry and Definiteness of a Regression Matrix for Demand Functions. UTTAM CHAND, University of North Carolina.

The importance of relations between two sets of variates (e.g. the study of relations of the prices to the quantities of several commodities) invariant under linear transformations of one set of variates contragredient to those of the other was first pointed out by Hotelling. In the study of related demand functions no suitable statistical tests have existed for testing the hypotheses of symmetry and negative definiteness of the regression matrix of prices on quantities. The test proposed here for the hypothesis of symmetry is exact and invariant under all contragredient transformations. A separate test studied for both symmetry and negative definiteness satisfies the property of invariance but its distribution depends on a nuisance parameter which is the non-zero root of a certain determinantal equation. The likelihood ratio criterion under the hypothesis of symmetry leads to a multilateral matric equation which represents $\frac{1}{2}p(p+1)$ equations of the third degree in $\frac{1}{2}p(p+1)$ unknown regression coefficients for the p-variate case, and does not admit of a unique solution.

26. The Distribution of Extreme Values in Samples whose Members are Stochastically Dependent. BENJAMIN EPSTEIN, Wayne University, Detroit.

In this paper the following problem is considered. To find the distribution of largest and smallest values in samples of size n drawn from a random process subject to the following conditions:

(i) observations $x_1, x_2, \cdots, x_n$ are taken in order from some random process.

(ii) the random process is such that successive observations x_i and x_{i+1} are jointly dependent. The joint distribution is described analytically, independently of i, by a two-dimensional d.f.

$$F_2(x, y) = \text{Prob}\,(x_i \leq x, x_{i+1} \leq y), \qquad 1 \leq i \leq n-1,$$

(iii) $F_2(x, y) = F_2(y, x)$

(iv) Any other pairs of observations (x_i, x_{i+j}), $1 \leq i \leq n-1$, $2 \leq j \leq n-1$, are assumed to be independent.

The results in this paper generalize the special situation where all observations are independent. More general cases than those covered by (i)–(iv) are briefly considered.

27. **On Age-Dependent Stochastic Branching Processes.** RICHARD BELLMAN AND THEODORE E. HARRIS, Stanford University and The RAND Corporation, Palo Alto and Santa Monica, California.

An initial particle has a random life length T with c.d f. $G(t)$. At death it is replaced by a random number N of similar particles; $P(N = n) = q_n$. Particles produced have the same distributions of life-length and replacement as the original one.

Let $z(t)$ = number of particles at time t, $h(s) = \sum_{n=0}^{\infty} q_n s^n$, $F(s, t) = \sum_{n=0}^{\infty} P(z(t) = n)s^n$. The integral equation $F(s, t) = \int_0^t h[F(s, t - y)]\, dG(y) + s[1 - G(t)]$ uniquely determines $F(s, t)$. When suitable restrictions are put on $h(s)$ and $G(t)$, results of Feller can be applied to study the asymptotic behavior of the moments of $z(t)$, which satisfy linear integral equations of the convolution type, and further special results on the moments can be obtained. The condition $\Sigma n q_n > 1$ and certain further restrictions insure that $z(t)/e^{bt}$ converges in probability as $t \to \infty$, where b is a certain constant The m.g.f. $\phi(s)$ of the limiting distribution satisfies the equation $\phi(s) = \int_0^{\infty} h[\phi(se^{-bv})]\, dG(y)$. Further restrictions imply that $\phi(s)$ is analytic in a neighborhood of $s = 0$, and that the corresponding distribution is absolutely continuous.

28. **Cuboidal Lattices.** G. S. WATSON, Institute of Statistics, North Carolina.

Yates has given two series of partially balanced incomplete block designs, square and cubic lattices, which enable the experimenter to test respectively k^2 and k^3 varieties in blocks of size k. Harshbarger has recently given a series of designs, rectangular lattices, which supplement Yates' square lattices.

In this paper two series of designs are given called cuboidal lattices, supplementing the cubic lattice series. They may be used to test respectively $k^2(k + 1)$ and $k(k + 1)^2$ varieties in blocks of k, when the number of reflications is a multiple of 3. Interblock information may be recovered. The first series has a relatively simple analysis and should prove useful.

This work was sponsored by the Office of Naval Research.

29. **Transformations Induced by Series Approximation of Prior Probability Amplitude.** ARCHIE BLAKE, Office of The Surgeon General, U. S. Army.

Consider a class A of mutually exclusive and exhaustive possible outcomes of a test. (We assume A finite; this condition can under suitable conditions be removed at a later stage by a limiting process.) For a hypothesis h, let u be the vector whose value, for each member a of A, is the square root of the prior probability of a and h jointly. This vector is called the probability amplitude; its norm, the scalar product $u'u$, is proportional to the prior probability of h, the constant of proportionality being determined by comparing the norms of the u's for all h. Let the test leave the alternatives of a subclass B of A still possible, while ruling out the members of $A - B$. Represent this test by a vector r having the value 1 on B and 0 on $A - B$. Define d on AA as a matrix equal to r on the main diagonal and zero elsewhere. The posterior probability is proportional to the form value $u'du$, the norm of the projection of u on a subspace determined by suppressing the coordinates of $A - B$. Consider the transformation $u = tv$, t being a matrix on AA and v a vector on A. Then $u'du$ takes the form $v't'dtv$. Denote $t'dt$ by e. If u is approximated as a partial sum of the series tv, i.e. by truncating v with a subclass C of A, the truncation induced on e is that with the minor on CC. (How much of the prior probability norm is

retained with a particular truncation is most easily seen if t is orthogonal, for then the transform of $u'u$ is $v'v$)

For example, in an agricultural experiment, let A be the composite of P, the class of plots, and Y, the class of possible yields on a plot Then u takes the form of a second order tensor or matrix on PY, while d and t are fourth order tensors. For some member y of Y, it often happens that some of the initially most probable, numerous, and economically consequential hypotheses will be such that for them the values of $u(y)$ are predominantly high on some row of plots, low on another row, etc. The transformation $u = tv$ on P induces the transformation $e = t'dt$; this is R. A. Fisher's transformation, performed, however, on d instead of on the yields themselves. The truncation of v and e corresponds to Fisher's relegating the higher interactions to error. This calculation may be accompanied by a linear transformation on Y, e.g. in series of orthogonal functions (Such series are not subject to the disadvantage of classical Gram-Charlier series, which are expressed in terms of the probability instead of its square root, that their partial sums can be at places negative)

30. On the Utilization of Marked Specimens in Estimating Populations of Flying Insects. CECIL C. CRAIG, University of Michigan, Ann Arbor.

The experimenter catches flying insects, say butterflies, marks and immediately releases them It is assumed that all the insects in a segregated area are equally liable to capture whether unmarked or marked, even several times, and that the population is stable for this period over which a series of captures is made. From the record of insects caught once, twice, three times, and so on, the problem is to estimate the total population. Two mathematical models which seem appropriate are considered and four methods of estimation are compared with respect to the large sample variances of the estimates they give.

31. On a Probability Distribution. MAX A. WOODBURY, University of Michigan, Ann Arbor.

In this paper the probability of x successes in n trials of an exent is computed for the case when the probability of success in a given trial depends only on the number of previous successes. The solution $P(n, x)$ satisfies the equation of partial differences

$$P(n+1, x+1) = (q - q_x)P(n, x) + q_{x+1}P(n, x+1)$$

in the case when $q = 1$. The boundary conditions are obviously $P(0,0) = 1$ and $P(n, x) = 0$ for $x < 0$ or $> n + x$ The solution of this equation is obtained by use of a generating function and $P(n, x)$ proves to be the xth term in the expansion of q^n by means of Newton's divided difference formula given the values $q_0^n, \cdots, q_x^n, \cdots, q_n^n$. Specifically, by setting $q = 1$, one obtains the result

$$P(n, x) = p_0 p_1 \cdots p_{x-1} \sum_{i=0}^{x} q_i^n / [(q_i - q_0)(q_i - q_1) \cdots (q_i - q_{i-1})(q_i - q_{i+1}) \cdots (q_i - q_x)].$$

In the case $p_x = p_0$ one has the result

$$P(n, x) = \frac{p_0^x}{x!} \frac{d^x}{dq^x} (q^n)\Big|_{q=q_0}$$

which yields the usual result on simplification.

32. Distribution-Free Tests of Data from Factorial Experiments. G. W. BROWN AND A. M MOOD, Iowa State College.

A device for avoiding the assumption of normality in analysis of variance problems was

developed by M. Friedman (*Am Stat Assoc. Jour.*, Vol. 32 (1937), pp 675–701) in which the values of the observations were replaced by their ranks

An alternative approach is presented here in which medians are used to construct certain contingency tables, and the various null hypotheses of interest are easily tested by means of the ordinary chi-square criterion applied to such tables. These tests·

(1) Avoid the assumption of normality.
(2) Are particularly sensitive to differences in locations of cell distributions but not to their shapes.
(3) Usually require very little arithmetic computation.

The tests and the relevant distribution theory have been worked out for some of the simpler experimental designs.

33. On Sums of Symmetrically Truncated Normal Random Variables.

FRED C. ANDREWS AND Z. W. BIRNBAUM, University of Washington, Seattle.

Let X_a be the random variable with the probability density

$$f_a(X) = Ce^{-X^2/2} \quad \text{for} \quad |X| \leq a, \qquad f_a(X) = 0 \quad \text{for} \quad |X| > a,$$

and let $S_a^{(n)} = \sum_{i=1}^{n} X_a^{(i)}$ where $X_a^{(1)}, \cdots, X_a^{(n)}$ are independent determinations of X_a. The problem considered is, for given n, $T > 0$, $\epsilon > 0$, determine a such that $P(|S_a^{(n)}| \geq T) = \epsilon$. The exact solution of this problem would require laborious computations. In this paper a method is given for obtaining approximate values of a which are "safe" i e such that $P(|S_a^{(n)}| \geq T) \leq \epsilon$.

34. On the Foundation of Statistics.

MAX A. WOODBURY, University of Michigan, Ann Arbor.

The results on this paper are part of the author's University of Michigan dissertation, "Probability and Expected Values" The work covered by this paper was sponsored by the Office of Naval Research One may take the notion of an expected value as the basis for the theory of Statistics; i e. a linear functional on a linear space of random variables (real valued functions defined over a population). The space is called statistical if it contains all constant functions and the expected value of such constant functions is just the constant and if the expected value of a non-negative function is non-negative A statistical space is called strong if it contains with a random variable also the random variable whose values are the absolute values of the given random variable. Every expected value defines a probability measure over a quorum of subsets of the population and it is shown that the integral of the random variable, if it exists, coincides with the expected value. Further it is shown that if the statistical space is strong the integral necessarily exists and also that a necessary and sufficient condition that the quorum be a field is that the statistical space be strong.

35. Finitely Additive Probability Functions.

MAX A. WOODBURY, University of Michigan, Ann Arbor.

The results in this paper are part of the work in the author's University of Michigan dissertation, "Probability and Expected Values." The work covered by this paper was sponsored by the Office of Naval Research A quorum is a family of sets that contains with each pair of disjoint sets also their union and also the complement of any of its sets. Trivially a quorum is required to contain at least one set and hence at least the universe set or population and the empty set. An extension of the notion of a finitely additive probability measure function to quorums is given and proved to be equivalent to the usual

definition in case the quorum is a field of sets. The extension of a quorum of sets relative to the probability measure function is investigated using the properties of the inner and outer measure. The upper and lower integrals are defined and a condition for the existence of the integral is given. When the quorum is a field it is shown that integrability of a function implies the existence of the distribution function. This last result is well known in the case where the probability measure function is completely additive.

36. **On Inverting a Matrix via the Gram-Schmidt Orthogonalization Process.** MAX A. WOODBURY, University of Michigan, Ann Arbor.

The application of the classical Gram-Schmidt orthogonalization process to the factorization of a correlation matrix is accomplished by considering the inner product $[x, y] = E(xy)$ in the linear space determined by the statistical variables $x_1, x_2, \cdots, x_n$. In this way a representation of the original set of statistical variables in terms of an orthonormal set is obtained. (By an orthonormal set we mean a set $\xi_1, \xi_2, \cdots, \xi_n$ such that $E(\xi_i \xi_j) = 0$ for $i \neq j$ and $E(\xi_i^2) = 1$.) The matrix of coefficients $B = (b_{ij})$, where $x_i = \sum_1^i b_{ij}\xi_j$, has the property that $C = BB'$ where $C = (E(x_i x_j))$ and $'$ denotes the transpose. Further the matrix B is triangular hence B^{-1} is readily computed, from which one obtains at once $C^{-1} = (B^{-1})'B^{-1}$. The quantities b_{ij} are readily obtainable by the method of determinants (Dwyer and Waugh, *Annals of Math. Stat.*, Vol. 16 (1945), pp. 259–271, cf. pg. 264) formerly called the method of multiplication and subtraction with division.

37. **Certain Properties of the Multiparameter Unbiased Estimates.** G. R. SETH, Iowa State College.

If $\theta^* = (\theta_1^*, \theta_2^*, \cdots, \theta_q^*)$ is an unbiased vector estimate of $\theta = (\theta_1, \theta_2, \cdots, \theta_q)$ in the density function $p(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_q)$ having the smallest concentration ellipsoid among the class of unbiased estimates of θ, and further if ϵ is any statistic of q components having $E(\epsilon) = 0$ and finite covariance matrix, then ϵ is uncorrelated with θ^*.

If a set of sufficient statistics $(T_1, T_2, \cdots, T_p)$, $p \leq q$, exists for estimating θ, then corresponding to any unbiased vector estimate ϕ^* of θ, there exists an unbiased estimate of θ depending on $T_1, T_2, \cdots, T_p$ alone, where the latter has a concentration ellipsoid equal to or contained in that of the former.

When $q = 1$, and ϕ^* has the smallest variance among the class c formed by unbiased estimates of θ which are functions of θ^* having a finite variance, and the set of polynomials with respect to the distribution function of ϕ^* is complete, then ϕ^* is the only element in the class c. For $q > 1$, the result holds when the "variance" is replaced by the "concentration ellipsoid."

38. **A Class of Lower Bounds for the Variance of Point Estimates.** DOUGLAS CHAPMAN, University of California, Berkeley.

A class of lower bounds for the variance of point estimates is derived by means of the calculus of finite differences under very weak restrictions and it is shown that they give valid lower bounds for certain parameter estimation problems for which the Cramér-Rao formula is invalid. In some cases even when the latter lower bound exists a sharper lower bound may be found in the class here defined. On the other hand when it exists, the Cramér-Rao lower bound is asymptotically superior to any of this class.

39. **Standard Errors and Tests of Significance for Interpolated Medians.** CHURCHILL EISENHART AND MIRIAM L. YEVICK, National Bureau of Standards.

If a sample of N observations is grouped by a sequence of class intervals with boundaries $-\infty, \cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots, +\infty$, where x_0 is the largest boundary point for which the observed 'fraction below', p_B, is less than $\frac{1}{2}$, and x_1 is the smallest boundary point for which the observed 'fraction above', p_A, is less than $\frac{1}{2}$, so that the observed 'central fraction', p_C, between x_0 and x_1 is positive, then, at least for the case of N large, standard textbooks take as the median of the grouped data the interpolated median,

$$m = x_0 + b(x_1 - x_0)$$

where

$$b = (\tfrac{1}{2} - p_B)/p_C .$$

The literature is silent regarding the sampling properties of such medians, and regarding tests of significance appropriate to them. Let P_B and P_C be the population fractions below x_0, and between x_0 and x_1, respectively, and let u and β be the population analogs of m and b obtained by replacing p_B and p_C in the above equations by P_B and P_C, respectively. It is shown that m is asymptotically normally distributed about u so defined with asymptotic variance given by

$$\frac{1}{NY_C^2}\left[P_B(1 - P_B) - 2\beta P_B P_C + \beta^2 P_C(1 - P_C)\right]$$

where

$Y_C = \dfrac{P_C}{x_1 - x_0}$ = ordinate of 'central rectangle' of 'population histogram'. The classical formula for the variance of a median can be obtained as the limit of the above when $(x_1 - x_0) \to 0$ *with* $P_B \to \frac{1}{2}$.

In addition, tests of hypotheses regarding the value of the 'interpolated median of the population', u, and regarding the difference, $u_2 - u_1$, of the interpolated medians of two populations, are developed (1) by utilizing the above asymptotic results, and (2) by utilizing the Neyman-Pearson likelihood-ratio-test approach.

40. Some Efficient Range-Estimates of Variation. NILAN NORRIS, Hunter College, New York.

The commonly used sample range (in the sense of the difference between the largest and smallest of the variates) is one of an unlimited number of range or difference-measures which can be used to scale parent populations For samples drawn from a Type III universe, the maximum-likelihood estimate of dispersion is given by $A - G$, where A is the sample arithmetic mean and G is the sample geometric mean. For samples drawn from a Type V universe, a 100% efficient estimate of absolute variation is given by $G - H$, where G is the sample geometric mean and H is the sample harmonic mean. Under certain general conditions usually fulfilled, the standard errors of both of these range-measures of absolute dispersion may be estimated from expressions obtained by application of the Laplace-Liapounoff theorem. The two parametric methods of estimating absolute variation as developed in this paper are likely to be most useful when the form of the parent universe is known, and it is either too expensive or impossible to obtain samples large enough to permit the use of inefficient estimates. An example of such a case is the learning curve encountered in the analysis of frequency of occurrence of aircraft accidents by hours of flying experience of pilots in training. E. J. G. Pitman, *Proc. Camb Phil. Soc.*, Vol 33 (1937), pp. 217–218, has discussed the scaling of the Type III distribution. The method of scaling given by Pitman differs from the method of estimation developed in this paper for the Type III universe.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Franz L. Alt has resigned his position with the Ballistic Research Laboratories at Aberdeen to join the National Bureau of Standards where he is in charge of the Computation Laboratory of the "National Applied Mathematics Laboratory."

Dr. Edward W. Barankin has been promoted to Assistant Professor and Research Associate at the Statistical Laboratory, University of California, Berkeley, California.

Dr. Stanley Clark has accepted an associate professorship of Education at the College of Education, University of Saskatchewan, Saskatoon, Canada.

Dr. Gerald J. Cox has resigned his position as Research Chemist in the Chemical Division of Corn Products Refining Co., Argo, Illinois to accept an appointment as Professor of Dental Research in the School of Dentistry of the University of Pittsburgh

Mr. S. Lee Crump has resigned his assistant professorship at Iowa State College to accept a position in the Atomic Energy Project, University of Rochester.

Dr John H. Curtiss, Chief of the National Applied Mathematics Laboratories of the National Bureau of Standards, has assumed temporary additional duties as Acting Chief of the Institute for Numerical Analysis. The Institute for Numerical Analysis, located on the U.C.L.A. campus, was established by the National Bureau of Standards with the support of the Office of Naval Research and the United States Air Force for the two-fold purpose of pursuing mathematical research aimed at the development of numerical techniques for the full exploitation of the newer large-scale electronic computing machines and for performing numerical computations basic to the extension of the frontiers of science.

Mr. Walter T. Federer has resigned his position at the Statistical Laboratory at the Iowa State College to accept a position as Professor of Biological Statistics in the Department of Plant Breeding at Cornell University.

Dr John Gurland, who received his Ph.D. in mathematical statistics from the University of California in August, 1948, is now a Benjamin Pierce Instructor in Mathematics at Harvard University.

Dr. Joseph L. Hodges, Jr. has been promoted to Instructor and Research Associate at the Statistical Laboratory, University of California, Berkeley.

Dr Cyril J. Hoyt has resigned his position as Research Associate with the Department of Education at the University of Chicago to accept an appointment as Associate Director of the Bureau of Educational Research, University of Minnesota.

Dr Tjalling C. Koopmans has been promoted to Professor of Economics at

the University of Chicago and also Director of Research of the Cowles Commission for Research in Economics

Dr Eugene Lukacs, formerly at Our Lady of Cincinnati College, has accepted a position as statistician at the United States Naval Ordnance Test Station at Inyokern, California

Mr. Frank Jones Massey, Jr., who has been in the Department of Mathematics at the University of Maryland, has accepted an assistant professorship in the Department of Mathematics at the University of Oregon, Eugene, Oregon.

Miss Judith Moss has resigned her position at the National Bureau of Economic Research and is now with the Port of New York Authority as an Economic Analyst in the Planning Bureau.

Dr Richard Otter has accepted an assistant professorship in the Department of Mathematics at the University of Notre Dame.

Dr. Nathan Grier Parke, III has been appointed Research Fellow of the Massachusetts General Hospital and Associated Research Director of the Harvard Piatric Study.

Dr. Joseph A. Pierce is now serving as Chairman of the Division of Natural Science and Mathematics at the Texas State University for Negroes, Houston 4, Texas.

Dr. Saul B. Sells, former Assistant to the President of the A. B. Frank Co. of San Antonio, Texas, has joined the staff of the Department of Psychology of the Air University, School of Aviation Medicine, Randolph Field, Texas.

Dr. Otis A Pope, who was with the Office of Foreign Agricultural Relations, U. S. Department of Agriculture, Technical Collaboration Branch, Washington, D. C., died September 28th, 1948.

Special Summer Session in Survey Research Techniques

The Survey Research Center of the University of Michigan will hold its special summer session in Survey Research Techniques from July 18 to August 13, 1949.

The following courses will be offered: Introduction to Survey Research, Survey Research Methods, Sampling Methods in Survey Research (introductory and advanced), Mathematics of Sampling, Statistical Methods in Survey Research, Techniques of Scaling.

In addition the introductory courses will be given from June 20 to July 16. This will permit students who are attending the full eight-week summer session of the University (June 20 to August 13) to register for the introductory courses during the first four weeks.

It is expected that this special session will attract men and women employed in market research or other statitical work and university instructors and graduate students with a particular interest in this area of social science research.

All courses are offered for graduate credit and students must be admitted by

the Graduate School. Inquiries should be addressed to the Survey Research Center, University of Michigan, Ann Arbor, Michigan.

Summer Courses in Statistics at Michigan

In addition to the special courses in Survey Research Techniques, the following courses of special interest to students of statistics are among those offered by the mathematics department of the University of Michigan in the Summer Session, June 20 to August 13: Finite Differences (Fischer), Probability (Copeland), Theory of Statistics I and II (Carver), Significance Tests (Dwyer), Computational Methods (Dwyer), Theory of Estimation and of Significance Tests (Craig) and Seminar (Craig).

The International Congress of Mathematicians

No summer meeting of the Institute of Mathematical Statistics is planned for 1950 because of the meeting of the International Congress of Mathematicians which will be held in Cambridge, Massachusetts August 30 to September 6, 1950. The following statement has been prepared by the organizing committee:

An International Congress of Mathematicians will be held in Cambridge, Massachusetts, in 1950 under the auspices of the American Mathematical Society. The Society originally planned to act as host for a Congress in September, 1940, which was also scheduled to meet in Cambridge. At the 1936 Congress in Oslo, Norway, the invitation for the 1940 Congress was issued by the American delegation in the name of the American Mathematical Society. Plans for the 1940 Congress were practically completed when the outbreak of World War II in September, 1939, made it necessary for the Society to postpone the Congress to a more favorable date. An Emergency Committee was established to carry on in the interim and, on recommendation of this Committee, the Council of the Society voted to hold the Congress in 1950.

The 1950 Congress will be the third International Congress of Mathematicians to be held on the continent of North America. The first was held at Northwestern University in 1893 and the second at the University of Toronto in 1924. International Congresses were held at intervals of approximately four years, except when war intervened, until 1936. There has been no international gathering of mathematicians since that time and it is the sincere hope of the Organizing Committee that the gathering in 1950 will be a truly international one, that the American mathematicians will attend in large numbers, and that all other countries will be well represented. The Council of the American Mathematical Society has voted unanimously to hold a Congress which will be open to mathematicians of all national and geographical groups.

Time and Place. The dates for the Congress have been fixed as August 30–September 6, 1950. Harvard University will be the principal host institution. A number of other institutions in metropolitan Boston will join in the entertainment of Congress visitors by arranging special features on their campuses.

Type of Congress. In recent years mathematicians have been much impressed by the success of the conference method for presenting recent research in fields where vigorous advances have just been made or are in progress. In view of the success of mathematical conferences on special topics which have been held in Russia, France and Switzerland and, more recently, at the Princeton Bicentennial Celebration, the 1950 Congress will include Conferences in several fields. For the 1940 Congress, Conferences in four fields had been planned. The number of Conferences was thus restricted lest the introduction of a promising and novel feature result in failure through the dissipation of interest and energy.

Following the established custom, the Organizing Committee plans to have a number of invited hour addresses by outstanding mathematicians. In addition, sectional meetings for the presentation of contributed papers not included in Conference programs will be held in the following fields: I, Algebra and Theory of Numbers; II, Analysis; III, Geometry and Topology; IV, Probability and Statistics, Actuarial Science, Economics; V, Mathematical Physics and Applied Mathematics; VI, Logic and Philosophy, VII, History and Education.

The official languages of the 1950 Congress will be English, French, German, Italian, and Russian.

Organization. The plans for the Congress are under the supervision of an Organizing Committee which was elected by the Council of the American Mathematical Society in February, 1948. The Chairman is Professor Garrett Birkhoff of Harvard University and the Vice Chairman is Professor W. T. Martin of Massachusetts Institute of Technology. Other members of the committee are: Professors J L. Doob, G. C. Evans, J. R. Kline, Solomon Lefschetz, Saunders MacLane, Dean R G. D. Richardson, Professors Oswald Veblen, J. L. Walsh, D. V. Widder, Norbert Wiener, and R. L. Wilder.

Many of the subventions promised for the 1940 Congress are still available. A Financial Committee under the chairmanship of Professor John von Neumann is endeavoring to secure additional funds. Besides support from Harvard University and Massachusetts Institute of Technology, generous subventions have been subscribed for the Congress by the Carnegie Corporation, the Institute for Advanced Study, the National Research Council, and the Rockefeller Foundation.

An Editorial Committee under the chairmanship of Professor Salomon Bochner will assume responsibility for the publication of the Proceedings of the Congress.

Professor J. R Kline of the University of Pennsylvania has been named Secretary of the Congress and Dr. R. P Boas, Executive Editor of Mathematical Reviews, has been designated Associate Secretary.

Entertainment. Harvard University has offered the use of its dormitories and dining rooms for mathematicians and their guests for the period of the Congress. The Organizing Committee hopes that it will be possible to furnish room and board without charge to all mathematicians from outside continental North America who are members of the Congress. Congress membership fees and rates for room and board will be announced well in advance of the opening of the Congress.

The Entertainment Committee, of which Professor L. H. Loomis of Harvard University is Chairman, is planning many interesting features, including a reception, garden party, symphony concert, and banquet. It is hoped that American mathematicians will be able to assist in the entertainment by putting their automobiles at the disposal of the Entertainment Committee for trips to be made out of Cambridge.

Every effort will be made to facilitate the travel at reasonable cost of foreign participants while in the United States. Previous to the Congress, opportunity will be given them to see New York City under the guidance of some mathematicians.

Information. Detailed information will be sent in due course to individual members of the American Mathematical Society and to foreign mathematical societies and academies. Others interested in receiving information may file their names in the office of the Society, and such persons will receive from time to time information regarding the program and arrangements.

Communications should be addressed to the American Mathematical Society, 531 West 116th Street, New York City 27, U. S. A.

New Members

The following persons have been elected to membership in the Institute

(August 16, 1948 to November 30, 1948)

Alman, John E., M.A. (Claremont Colleges) Instructor in Mathematics, College of Liberal Arts, Boston University, *216 Gardner Road, Brookline 46, Massachusetts.*

Andrian, Jane F., M.S. (Western Reserve Univ.) Graduate student at University of California, *1222 C Ashby Avenue, Berkeley 2, California.*

Arbuckle, Richard A., B.S. (Baldwin Wallace College) Research-Industrial Fellow at Purdue University, *F Ph.A. 530-3 Airport Road, Lafayette, Indiana.*

Barankin, Edward W., Ph.D. (Univ. of Calif.) Assistant Professor of Mathematics and Research Associate in Statistical Laboratory, University of California, Berkeley, California

Blum, Julius R., Student in mathematical statistics at the University of California, *1957 Acton Street, Berkeley 2, California.*

Bronfenbrenner, Mrs. Jean, M.A. (Univ. of Chicago) Research Assistant, Cowles Commission, University of Chicago, Chicago 37, Illinois

Burns, Loren V., B.S. (Washburn College, Topeka, Kansas) Technical Director, MFA Milling Co., *Box 1585 S.S.S., Springfield, Missouri.*

Clement, Edwin G., M.B.A. (Univ. of Chicago) Captain, Chief of Management Control Branch, Headquarters, Strategic Air Command, Andrews Air Force Base, Washington 20, D. C.

Cramer, George F., Ph.D. (Univ. of Missouri) Mathematician, U. S. Navy Department, Washington, D. C., *112 Quincy Street, Chevy Chase 15, Maryland.*

Degan, James W., A.B. (Univ. of Chicago) Research Assistant, Psychometric Laboratory, University of Chicago, *1128 East 61st Street, Chicago 37, Illinois.*

Dodd, Stuart C., Ph.D. (Princeton) Research Professor of Sociology and Director of Public Opinion Laboratory, *4725—45th Avenue, N.E., Seattle 5, Washington.*

Donnelly, Tom G., M.A. (Queen's Univ.) Graduate student at the University of North Carolina, *Room 213 "B", Chapel Hill, North Carolina*

Edwards-Davies, Harold D, Special Lecturer, Department of Mathematics, Dalhousie University, *67 Seymour Street, Halifax, N.S., Canada.*

Ellner, Henry, Ch E. (College of City of New York) Statistician (Physical Sciences) *1-C Oak Grove Drive, Baltimore 20, Maryland.*

Feigenbaum, Armand V., M S. (Mass Institute of Tech.) General Electric Company, Room 257, Building 23, Schenectady, New York

Festinger, Leon, Ph D (Univ. of Iowa) Assistant Professor of Psychology, Research Center for Group Dynamics, University of Michigan, Ann Arbor, Michigan.

Frame, James S., Ph D (Harvard) Professor and Head of Department of Mathematics, Michigan State College, Lansing, Michigan.

French, Benjamin J., M.Ed (Univ of New Hampshire) Examiner, Educational Testing Service, *Matthews Road, Keene, New Hampshire.*

Gaffey, William R., A B. (Univ. of Calif.) Research Assistant, University of California, *2306 Grant Street, Berkeley 4, California*

Goodman, Leo A., A B (Syracuse University) Research Assistant in Mathematical Statistics and Graduate student at Princeton University, Fine Hall, Princeton University, Princeton, New Jersey

Hader, Robert J., Ph D (North Carolina State College) Instructor and Research Assistant, Institute of Statistics, North Carolina State College, Raleigh, North Carolina.

Haley, Kenneth D., M S (Stanford Univ) Assistant Professor of Mathematics, Acadia University, Wolfville, Novia Scotia, Canada

Kahn, Louis B., M S (Univ. of Wisconsin) Research Associate, University of Wisconsin, *Box 16-F, Badger, Wisconsin.*

Katz, Irving, B.S. (College of City of New York) Statistician, Strategic Air Command, *379—87 Place, S E., Washington 19, D. C.*

Kientzle, Mary J., Ph D. (Univ of Ill) Assistant Professor of Psychology, Department of Psychology, Washington State College, Pullman, Washington

Koditschek, Paul, Ll. D (Univ. of Vienna) Research Associate, Scientific Research Service, Columbia University, *319 W. 13th Street, New York 14, New York.*

Levin, Howard S., S B. (Univ. of Chicago) Electronic Engineer, Glenn L. Martin Co., *532 Addison Street, Chicago 13, Illinois*

Levine, George J., B.S. (Brooklyn College) Actuarial Mathematician, *5109—1st Street, North, Arlington, Virginia.*

Liverman, J. G., B.A. (Cantab) Civil Servant, Ministry of Fuel and Power, *21 Ascot Court, Grove End Road, London, N.W. 8, England*

Loeve, Michel, Ph.D. (Sorbonne, Paris) Professor and Research Associate in Statistical Laboratory, Durant Hall, University of California, Berkeley, California

Loo, Ching-Tsu, Ph.D. (Univ of Chicago) Research Associate, Statistical Laboratory, University of California, Berkeley, California

Lubin, Ardie, B S. (Univ. of Chicago) Statistician, Psychology Department, Maudsley Hospital, Denmark Hill, S.E. 5, London, England.

Moses, Lincoln E., A.B. (Standord Univ.) *7 Perry Lane, Menlo Park, California.*

Mourier, Edith, Licencie-ès-sciences (Univ. of Caen, France) Teaching Assistant, Statistical Laboratory, University of California, Berkeley, California.

Osborne, Ernest L., L L B (LaSalle Univ) Economic Analyst, Department of the Army, *Chancery Apartments, 3130 Wisconsin Avenue, N.W., Washington 16, D. C.*

Pabst, William R. Jr., Ph.D (Columbia Univ.) Quality Control Division, Bureau of Ordnance, Navy Department, *3420 Quebec Street, N.W., Washington 16, D. C.*

Plackett, Robin L., M A (Cambridge, England) Lecturer in Mathematical Statistics, Department of Applied Mathematics, The University, Liverpool 3, England.

Proschan, Frank, M.A. (George Washington Univ.) Research Analyst, *1627 R. St , N.W., Washington 9, D. C*

Rau, A. Ananthapadmanabha, M S. (Iowa State College) Statistician and Agricultural Meteorologist, Department of Agriculture, Bangalore, Mysore State, India

Rees, Mina, Ph D. (Univ of Chicago) Head, Mathematics Branch, Office of Naval Research, R2719, T-3 Building, Washington 25, D. C.

Roberts, Spencer W. Jr., M.S. (Univ. of Michigan) Research Associate, University of Michigan Department of Engineering Research, *306 Thompson Street, Ann Arbor, Michigan*

Sarma, S. C., M Sc (Calcutta Univ) Graduate student in mathematical statistics at Columbia University, *1120 John Jay Hall, Columbia University, New York 27, New York.*

Schnelderman, Marvin A., B.S (College of City of New York) Statistician, Biological, National Institute of Health, T-6, 2215, Bethesda, Maryland.

Schull, William J., Ph.D. (Ohio State Univ.) Student at Ohio State University, Department of Zoology, Ohio State University, Columbia 10, Ohio.

Schweid, Samuel, B.S S. (College of City of New York) Statistician, Industry Division, Bureau of the Census, *1110 Monroe Street, N.W., Washington 10, D. C.*

Wallace, David L., B.S. (Carnegie Institute of Tech.) Graduate Student and Teaching Assistant in Mathematics, Carnegie Institute of Technology, *123 Lawrence Avenue, Homestead Park, Pennsylvania.*

Williams, Evan James, B.C. (Univ. of Tasmania) Research Officer, Section of Mathematical Statistics, Division of Forest Products, C S.I.R., P O. Box 18, South Melbourne, S.C 4, Australia

Zavrotsky, Andres, Head of the Statistical Department of the Venezuela Office for Social Insurance, *Mercedes a Luneta 39, Caracas.*

Correction of New Members in June, 1948 Issue:

Loizelier, Enrique Blanco, should be written as follows:

Blanco Loizelier, Enrique. (Ph D.) Professor of Statistics, Economics Faculty, Madrid University, Spain, *Nervion No. 4, Madrid, Spain.*

ELECTION OF OFFICERS AND COUNCIL AND REVISION OF BY-LAWS

At the membership meeting held at Cleveland on December 28, the following officers and members of the Council were elected:

President:	J. Neyman
President-Elect:	J. L. Doob
Council:	
3-year term	W. G. Cochran
	C. Eisenhart
	H. Hotelling
	A. Wald
2-year term	W. Feller
	P. G. Hoel
	H. Scheffé
	J. Wolfowitz
1-year term	Gertrude Cox
	M. A. Girshick
	J W. Tukey
	J. von Neumann

The By-Laws were also revised and further action was taken. More detailed accounts of this meeting will be sent directly to the members.

PAUL S. DWYER
Secretary

REPORT ON THE SEATTLE MEETING OF THE INSTITUTE

The thirty-sixth meeting and fourth Regional West Coast meeting of the Institute of Mathematical Statistics was held in Seattle, Washington, November 26–27, 1948. The sessions of November 27, 1948 were held jointly with the Biometric Society (Western N. A. Region). The meeting was attended by 91 persons, including the following 22 members of the Institute:

F. C Andrews, E W Barankin, Z. W. Birnbaum, A. H. Bowker, D G. Chapman, R. C. Davis, W. J Dixon, E. Fay, M. A. Girshick, P. Horst, H. M. Hughes, J. C. R. Li, F. Massey, J. Neyman, E Paulson, Elizabeth L Scott, Esther Seiden, M. Sobel, Z. Szatrowski, J. R. Vatnsdal, J E. Walsh and Zivia S. Wurtele.

At the morning session on November 26, Professor R. M. Winger of the University of Washington as chairman welcomed those attending the meetings, and the following program of contributed papers was presented:

1. *Estimation of the Variance of the Bivariate Normal Distribution.*
   Harry M Hughes, University of California
2. *Derivation of a Broad Class of Consistent Estimates*
   R C Davis, NOTS, Inyokern, California.
3. *Locally Best Unbiased Estimates.*
   Edward W. Barankin, University of California.
4. *Some Problems Related to the Distribution of a Random Number of Random Variables.*
   Edward Paulson, University of Washington.
5. *Asymptotic Expansions for the Distribution of Certain Likelihood Ratio Statistics.*
   Albert H. Bowker, Stanford University
6. *On a Problem of Confounding in Symmetrical Factorial Design.*
   Esther Seiden, University of California.
7. *Some Bounded Significance Level Tests of Whether the Largest Observations of a Set are Too Small.*
   John E. Walsh, Project RAND, Douglas Aircraft Corp., Santa Monica, Calif

The afternoon session of November 26, under the chairmanship of Professor J. Neyman of the University of California at Berkeley, had the following program:

1. Invited paper:
   Multiple Decision Functions.
   M. A. Girshick, Stanford University.
   Contributed papers:
2. *Determination of Optimal Test Length to Maximize the Multiple Correlation Coefficient.*
   Paul Horst, University of Washington.
3. *Some Numerical Comparisons of a Non-Parametric Test with Other Tests.*
   F. J. Massey, University of Oregon.

4. *On the Deviation of Extreme Values*
W J Dixon, University of Oregon
5 *The Optimum Size of Interval for Making Measurements of a Rocket's Angular Velocity.*
Edward A Fay, University of California.
6. *Stationary Time Series Analysis and Common Stock Price Forecasting.*
Zenon Szatrowski, University of Oregon.

At the morning session of November 27, with Professor W. F. Thompson of the University of Washington as chairman, the program consisted of the following papers:

1 Invited paper·
On the Place of Statistics in Fishery Biology
Willis S Rich, Stanford University and U S Fish and Wildlife Service.
Contributed papers.
2. *Distribution of the Number of Schools of Fish Caught per Boat.*
J Neyman, University of California.
3 *Some Problems in Fishery Research to which Statistical Methods are Applicable.*
Ralph Silliman, U. S. Fish and Wildlife Service, Seattle, Washington.
4. *The Application of the Hypergeometric Distribution to Problems of Estimating and Comparing Zoological Population Sizes.*
Douglas Chapman, University of California.
5 *Extension to Multivariate Case of Neyman's Smooth Test.*
Elizabeth L Scott, University of California.
6. *A Mathematical Theory of Vitamin A Metabolism in Fish.*
Norman E Cooke, Pacific Fisheries Experimental Station, Vancouver, B.C.

The afternoon session of November 27 was held under the chairmanship of Professor F. W. Weymouth of Stanford University, with the following program:

1. Invited paper.
Statistical Problem of Enumeration of Fish Eggs in the Sea.
Oscar E. Sette, U. S. Fish and Wildlife Service, San Francisco.
Contributed papers
2. *The Interactance Hypothesis*
Stuart C. Dodd, University of Washington.
3 *The Employment of Marked Members in Estimation of Animal Populations.*
Milner E. Schaefer, Stanford University
4 *Non-Response and Repeated Call-Backs in Opinion Polls.*
Z W Birnbaum, University of Washington.
5 *Statistical Problems Relating to Fisheries.*
J L. Hart, Pacific Biological Station, Nanaimo, B. C.

On November 26, at 6:30 o'clock there was a dinner for members and guests at the Edmond Meany Hotel.

Z. W. Birnbaum

REPORT ON THE CLEVELAND MEETING OF THE INSTITUTE

The Eleventh Annual Meeting of the Institute of Mathematical Statistics was held at the Statler Hotel, Cleveland, Ohio, on December 27–30, 1948. The

meeting was held in conjunction with the Annual Meeting of the American Statistical Association. The following 176 members of the Institute were in attendance:

P. H. Anderson, R L. Anderson, L. W. Anderson, Max Astrachan, G J. Auner, T A. Bancroft, B. Geoffrey, Z. W. Birnbaum, Archie Blake, E. E. Blanche, C I. Bliss, Dorothy S. Brady, A E. Brandt, G. W. Brown, T. H. Brown, M. A Brumbaugh, P. T. Bruybre, R. W. Burgess, I.W. Burr, J. M. Cameron, A. G Carlton, Harry Carver, F. R. Cella, Uttam Chand, R A. Chapman, Edmund Churchill, Herman Chernoff, W. G. Cochran, Jerome Cornfield, J. H. Cover, Gertrude M. Cox, C. C Craig, S. L. Crump, J. H Curtiss, D. A. Darling, W. L. Deemer, D. B DeLury, W E. Deming, Philip Desind, H. F. Dorn, C. W. Dunnett, P. S. Dwyer, Churchill Eisenhart, Benjamin Epstein, C. D. Ferris, Leon Festinger, C. H. Fischer, J. C Flanagan, M M. Flood, L R. Frankel, D A. S. Fraser, H. A. Freeman, Milton Friedman, H. C Fryer, E. F. Gardner, R. S Gardner, H. H. Germond, William Gomberg, E. L. Green, S. W. Greenhouse, J. Gurland, R. J Hader, K. W. Halbert, H J. Hand, M. H. Hansen, T. E. Harris, Boyd Harshbarger, P. M. Houser, J F Hofmann, Harold Hotelling, A. S. Householder, E E Houseman, Helen M Humes, C. C. Hurd, C. M Jaeger, R. J. Jessen, H. L Jones, Irving Katz, Leo Katz, Harriet J. Kelly, O. Kempthorne, A. W. Kimball, Jr., A. J. King, Leslie Kish, L. A Knowler, Lila F. Knudsen, C. F. Kossack, O. E. Lancaster, Marvin Lavin, S. B Littauer, Irving Lorge, F. W. Lott, Jr., Eugene Lukacs, P. J. McCarthy, C. J. Maloney, John Mandel, Nathan Mantel, H. B. Mann, E. S. Marks, Margaret Merrell, Helen Michaels, E B. Mode, A M Mood, Nathan Morrison, Dorothy J. Morrow, J. W. Morse, J. E. Morton, Jack Moshman, Frederick Mosteller, B. D. Mudgett, Hugo Muench, M R. Neifeld, R. H. Noel, G. E. Noether, J. I. Northam, H. W Norton, J. A. Norton, Jr , E. G. Olds, P. S. Olmstead, Bernard Ostle, A. E. Paull, Paul Peach, M. P. Peisakoff, E W. Pike, E. J. G Pitman, R. A. Porter, J. A. Rafferty, L. J. Reed, Olav Reiersol, William Reitz, F. D. Rigby, A. C Rosander, Herman Rubin, Erik Ruist, P. J. Rulon, Max Sasuly, F. E. Satterthwaite, L. J. Savage, Mary Ann Savas, Marvin Schniederman, Elizabeth Scott, G. R. Seth, Jack Sherman, S. S. Shrikhande, C. R. Simms, J. H Smith, G. W. Snedecor, Mortimer Spiegelman, B. R. Stauber, F. F. Stephan, Joseph Steinberg, J. V. Sturtevant, B. J. Tepping, W. R. Thompson, J W Tukey, Jan Vchytil, W. R. Van Voorhis, D. F. Votaw, Jr , F. M. Wadley, Helen M. Walker, D. L. Wallace, W. A. Wallis, G. S. Watson, Leonel Weiss, Samuel Weiss, E L Welker, M. E. Wescott, Phillips Whidder, D. R. Whitney, S. S Wilks, C P. Winsor, Gerald Winston, M A Woodbury, T. D. Woolsey, Holbrook Working, W. J. Youden

The first session, a joint session with the American Statistical Association, was held at 2:00 P.M. on Monday, December 27, at which time a paper entitled *Statistical Concepts in an Infinite Number of Dimensions* was presented by Professor David H Blackwell of Howard University. Professor E. J. G. Pitman of the University of Tasmania was chairman.

The second session of the opening day was devoted to contributed papers in mathematical statistics, and was held at 4:00 P.M. in conjunction with the American Statistical Association. Professor W. R. Van Voorhis of Fenn College was chairman. The following papers were presented:

1. *A Necessary Condition for a Certain Class of Characteristic Functions.* Preliminary report. Eugene Lukacs, NOTS, Inyokern, California and Our Lady of Cincinnati College, Cincinnati, Ohio
2. *Precision of Estimates from Samples Selected under Marginal Restrictions.* Preliminary report. Clifford J. Maloney, Research and Development Department, Camp Detrick, Frederick, Maryland.

3. *Properties of Maximum and Quasi-Maximum Likelihood Estimates of Parameters of a System of Linear Stochastic Difference Equations with Serially Correlated Disturbances.*
   Preliminary report Herman Rubin, Cowles Commission, University of Chicago.
4. *The Computation of Maximum Likelihood Estimates of Parameters of a System of Linear Stochastic Difference Equations with Serially Correlated Disturbances.*
   Herman Chernoff, Cowles Commission, University of Chicago
5. *Test Criteria for Hypotheses of Symmetry and Definiteness of a Regression Matrix for Demand Functions.*
   Uttam Chand, University of North Carolina.
6. *The Distribution of Extreme Values in Samples whose Members are Stochastically Dependent.*
   Benjamin Epstein, Wayne University.

A session on *Teaching Statistical Quality Control* was held on Monday evening, December 27, jointly with the Ohio Section of the American Society for Quality Control and Section on Training of Statisticians of the American Statistical Association. Professor Samuel S. Wilks of Princeton University presided at the session. The following two papers were presented.:

1. *Teaching Statistical Quality Control for Town and Gown.*
   Lloyd A Knowler, State University of Iowa.
2. *Instructional Aids for Statistical Quality Control.*
   Edwin G Olds, Carnegie Institute of Technology.

The session concluded with discussion by Professor Irving W. Burr of Purdue University, and Professor Theodore H. Brown of Harvard University.

A session on *Review of Statistical Methodology* was held jointly with the American Statistical Association at 2:00 P.M., December 28. Professor Frederick Mosteller of Harvard University presided. The following papers were presented:

1. *Surveys and Sampling.*
   Philip J McCarthy, Cornell University.
2. *Industrial Applications*
   Paul S Olmstead, Bell Telephone Laboratories.
3. *Biology, Physical Sciences and Experimental Design*
   W. J. Youden, National Bureau of Standards.

At 4:00 P.M. on Tuesday, December 28, Professor H. C. Fryer of Kansas State College presided at a joint session with the Biometric Society and Biometrics Section of the American Statistical Association. Papers presented were:

1. *Evaluation of Field Insecticides from Count of Survivors.*
   C I Bliss and Neely Turner, Connecticut Agricultural Experiment Station.
2. *Curved Dosage-Response Curves.*
   Oscar Kempthorne, Iowa State College.
3. *Statistical Variations in Contents of Dry-Filled Ampuls in Current Pharmaceutical Practice*
   M. W. Green, American Pharmaceutical Association, and Lila F. Knudsen, Food and Drug Administration.
4. *A Practical Method for Determining the Mean and Standard Deviation of Truncated Normal Distributions.*
   J Ipsen, Yale University.

The session was concluded with discussion by D. B. DeLury, Ontario Research Foundation; Lloyd Miller, Sterling-Winthrop Research Institute; C. Eisenhart, National Bureau of Standards; J. L. Northam, Kansas State College.

On Wednesday, December 29, at 2:00 P.M., Dr. W. Edwards Deming presided at a session on *Effects of Error in the Independent Variate in Regression Problems.* This meeting was held in conjunction with the Biometric Society and Biometric Section of the American Statistical Association. Papers presented were:

1. *Are There Two Regressions?*
   Joseph Berkson, Mayo Clinic.
2. *Present Status of the Theory.*
   Jerzy Neyman, University of California.
3. *The Identifiability of a Linear Relationship Between Variables which are Subject to Error*
   Olav Reiersol, Purdue University.

These papers were followed by discussion by Professor Churchill Eisenhart, National Bureau of Standards, Elizabeth L. Scott, University of California, and C. P. Winsor, Johns Hopkins University.

Professor Boyd Harshbarger, of the Virginia Polytechnic Institute, presided at the Wednesday afternoon session on contributed papers in mathematical statistics. Papers presented were:

1. *On Age-Dependent Stochastic Branching Processes.*
   Richard Bellman and Theodore E Harris, Stanford University, Palo Alto, California and the Rand Corporation, Santa Monica, California.
2. *Cuboidal Lattices.*
   G. S. Watson, Institute of Statistics, University of North Carolina.
3. *Transformations Induced by Series Approximation of Prior Probability Amplitude.*
   Archie Blake, Office of the Surgeon General, U. S. Army.
4. *On the Utilization of Market Specimens in Estimating Populations of Flying Insects.*
   Cecil C. Craig, University of Michigan.
5. *On a Probability Distribution.*
   Max A Woodbury, University of Michigan.
6. *Distribution-Free Tests of Data from Factorial Experiments.*
   G W. Brown and A. M. Mood, Iowa State College.
7. *On Sums of Symmetrically Truncated Normal Random Variables.*
   Fred C. Andrews and Z W. Birnbaum, University of Washington.
8. *On the Foundation of Statistics.*
   (By title). Max A. Woodbury, University of Michigan.
9. *Finitely Additive Probability Functions.*
   (By title). Max A. Woodbury, University of Michigan.
10. *On Inverting a Matrix via the Gram-Schmidt Orthogonalization Process.*
    (By title). Max A. Woodbury, University of Michigan.
11. *Certain Properties of the Multiparameter Unbiased Estimates.* Preliminary report.
    (By title). Gobind R. Seth, Iowa State College.
12. *A Class of Lower Bounds for the Variance of Point Estimates.*
    (By title). Douglas Chapman, University of California.
13. *Standard Errors and Tests of Significance for Interpolated Medians.*
    (By title). Churchill Eisenhart and Miriam L. Yovick, National Bureau of Standards.

A symposium on *Randomness and its Testing* occupied the 4:00 P.M. session on Wednesday. Dr. Walter A. Shewhart of the Bell Telephone Laboratories presided and the following papers were presented:

1. *Survey of Available Tests for Randomness.*
   W. Allen Wallis, University of Chicago.
2. *Power Functions of Tests for Randomness.*
   H B. Mann, Ohio State University.
3. *Power Functions of Non-Parametric Tests.*
   Ransom Whitney, Ohio State University.

Discussion was led by Bernice Brown, The Rand Corporation; Paul S. Olmstead, Bell Telephone Laboratories; E. J. G. Pitman, University of Tasmania.

The morning session on Thursday, December 30, was a joint session with the American Statistical Association, with Professor Jerzy Neyman of the University of California presiding The following two papers were presented upon invitation of the Institute:

1. *Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions.*
   T. W. Anderson, Columbia University.
2. *Some Aspects of the Theory of Testing Composite Hypotheses.*
   E L Lehmann, University of California

The Business Meeting was held at 10:00 A.M. on Tuesday, December 28. Dr Churchill Eisenhart presided. A report of this meeting is found elsewhere in this issue.

W. R. Van Voorhis
Assistant Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1948

The last few years have seen a considerable growth of the Institute. The upward trend has continued throughout 1948. The Institute has acquired 126 new members during the year, but this gain is to be balanced against losses due to resignation and suspension for non-payment of dues. The Institute starts the year 1949 with a membership of about 1,100 as against the membership of 1,037 at the beginning of 1948. While the net gain is still substantial, it is not quite as much as hoped for, and this may serve as an incentive for an increased membership drive in 1949. The constantly increasing interest and research activities in statistical theory and methodology are well reflected in our meetings and the publications appearing in the *Annals*.

Meetings. The growth of the Institute in the past few years has brought about a considerable increase in its various activities. This manifested itself particularly in the extensive and rich programs of the meetings held during the year 1948. In addition to the usual invited addresses and contributed papers, the programs included a considerable number of symposia on various important

subjects such as the theory of games (Berkeley, June; Madison, September), stochastic difference equations (Madison, September), scales of measurement (New York, April), sampling for industrial use (Berkeley, June), etc. The eleventh summer meeting was held in conjunction with the meetings of the American Mathematical Society and the Econometric Society (Madison, September). The eleventh annual meeting (Cleveland, December) was held in conjunction with the American Statistical Association, Econometric Society and Biometric Society. There were also three regional meetings: New York (April), Berkeley (June) and Seattle (November). The Berkeley meeting was held in conjunction with the Pacific Division of the American Association for the Advancement of Science and some of the sessions of the Seattle meeting were sponsored jointly with the Biometric Society.

To facilitate the organization of meetings and arrangements of programs, instead of a single program committee there were three program committees appointed, one for Eastern, one for Mid-Western and one for Far-Western meetings. These committees consisted of the following members. Eastern Committee: W. G. Cochran, C. Eisenhart (Chairman), F. Mosteller, and J Wolfowitz, Mid-Western Committee: C. C. Craig, H. B Mann, and A. M. Mood (Chairman); Far Western Committee: Z. W. Birnbaum, M. A Girshick, P. G. Hoel, and J. Neyman (Chairman). To coordinate the work of these three program committees, a coordinating committee was appointed consisting of J. W. Tukey (Chairman) and the three chairman of the three program committees. This committee was also charged with the responsibility of making recommendations to the Board of Directors as to times and places for future meetings. Another innovation introduced during the past year was the appointment of assistant secretaries in connection with the meetings. S. B. Littauer acted as assistant secretary for the New York meeting, K. J. Arnold for the summer meeting in Madison, Z. W. Birnbaum for the Seattle meeting and W. R. Van Voorhis for the Cleveland meeting The assistant secretaries were charged with the task of looking after the local arrangements that had to be made in connection with the meetings. The appointment of assistant secretaries proved to be a great success not only in facilitating the necessary local arrangements for meetings but also in relieving the burden on the secretary's office. On the basis of this year's experience, it seems very desirable to continue with this practice in the future.

No Rietz Memorial lecture was given in 1948 in accordance with a decision of the Board of Directors that these lectures should not be given every year. It is planned, however, to have a Rietz lecture for 1949 and the Board of Directors invited J. Neyman to deliver it.

The New Constitution. One of the major events of the year was the adoption of the new constitution at the meeting in Madison. The growth of the Institute in recent years made parts of the old constitution obsolete and the need for a revision was apparent. Our thanks are due to the Committee on Planning and Development which has devoted much time and consideration to the study of

the problem and prepared a draft of a revised constitution. M. H. Hansen was chairman of this Committee. Other members were: J. H. Curtiss, W. G. Cochran, W. Feller, J. Neyman, H. W. Norton, F. F. Stephan, J. W. Tukey, and W. A Wallis. A draft of the new By-Laws was prepared by J. W. Tukey, who acted as a subcommittee of the Committee on Planning and Development.

Annals. The growth of the Institute during the past few years has manifested itself also in a constantly increasing number of manuscripts submitted for publication in the *Annals* While it is very gratifying to see this upward trend, it raises some problems of financial nature. At the rate manuscripts are coming in, an expansion of the publication facilities of the Institute would seem very desirable. Increase of the volume of the *Annals* would, however, mean increased cost and the present financial situation of the Institute could not allow such an additional burden unless some new sources of income can be found. Apart from a possible increase in the cost of printing the *Annals*, it seems that additional expenditures will be necessary for secretarial help in 1949. It was decided at the membership meeting in Madison that additional funds be raised through the contributions of universities and other organizations with strong interest in mathematical statistics and through the contributions of the members. Appeals for such contributions were sent out and it is hoped that there will be a generous response.

The new constitution permits the appointment of responsible Associate Editors. This brings up the whole question of editorial set-up and policies. A committee with S. S Wilks as chairman was appointed to make a thorough study of the Institute's publication experience and to make recommendations as to publication policies and editorial set-up. Other members of this committee are: W G. Cochran, W. Feller, M. A. Girshick, J. Neyman, P. S. Olmstead, W. A. Wallis and J Wolfowitz. The committee gave much thought and consideration to the problems involved and will report to the newly elected officers and Council.

The *Annals* has developed under the leadership of the Editor, S. S. Wilks, to one of the outstanding professional journals. I am sure that I can speak for all our members in expressing the Institute's indebtedness to S. S. Wilks for his untiring and most successful work.

Committees. The problem of classification of statisticians in the Government service is naturally of considerable importance to the statistical profession. A committee consisting of W. E. Deming (chairman) and C. Eisenhart was appointed to make a thorough study of this question with a view to advising the Civil Service Commission. The committee prepared a report in which three main categories of statisticians in Government Service are distinguished: mathematical statisticians, statistical analysts and data-collecting statisticians. The report was transmitted to the Civil Service Commission with the approval of the Board of Directors. The members of this committee are to be commended for the excellent work they have done in spite of the severe limitation of time allotted by the Civil Service Commission. The work on the problem of classifica-

tion of statisticians still goes on and a committee of experts consisting of members of the Washington Statistical Society, the Institute of Mathematical Statistics, and the American Statistical Association has been set up to advise the Civil Service Commission on this problem. Our representatives on this committee of experts are: W. E. Deming, C. Eisenhart, M. H. Hansen and S. Weiss.

The advances in numerical computations in recent years has made an enlargement and reorganization of the Committee on Tabulation necessary. Its present members are: R. L. Anderson, C. Eisenhart (Chairman), A. M. Mood, F. Mosteller, H. G Romig, L E. Simon, and J. W. Tukey. The objectives of this committee, as outlined by the chairman are: (1) to prepare a comprehensive list of new mathematical tables that would be of value in statistical theory and applications, (2) to assemble an American Collection of "Tables for Statisticians", (3) to prepare a list of mathematical tables of importance in statistical theory and applications to be recommended for inclusion in the proposed National Bureau of Standards volume of "Tables for the Occasional Computer". To implement the program of the committee, the following sub-committees have been constituted: (1) "On Computing Centers" with L. E. Simon as Chairman, (2) "On Ranks and Runs" with A. M Mood as Chairman, (3) "On Serial Correlations" with R. L. Anderson as Chairman, (4) "On 2 x 2 Tables" with C. Eisenhart as Chairman, (5) "On Order Statistics" with F. Mosteller as Chairman, (6) "On Binomial, Poisson, and Hypergeometric Distributions" with H. G. Romig as Chairman, (7) "On Miscellaneous Tables" with J. W. Tukey as Chairman.

On the recommendation of the membership committee, consisting of H. Scheffé (chairman), C. C. Craig, P. G. Hoel and F. F. Stephan, the following members have been elected as Fellows· J. Berkson, E. L. Lehmann, E. J G. Pitman, H. E. Robbins and C. M. Stein. The members of the finance committee for 1948 were P. S. Dwyer (chairman), C F. Roos, L. A. Knowles and T. N. E Greville.

The Nominating Committee for 1948 consisted of W. Bartky (chairman), C. C. Craig, J. F. Daly, H. A. Freeman, E L. Lehmann and W. G. Madow. The committee nominated J. Neyman for President, J. L. Dobb for President-Elect and 24 Council members for the 12 positions to be filled. In accordance with the provisions of the new constitution, the Nominating Committee for 1949 has also been appointed. The members of this Committee are: W. G. Cochran (Chairman), M. H. Hansen, H. B. Mann, A M. Mood and H. G. Romig.

The Board of Directors has been exploring the possibilities for a closer cooperation with our colleagues abroad and for making foreign statistical publications more easily accessible to our members. In particular, there has been correspondance with Professor E. S. Pearson, Managing Editor of *Biometrika*, on the question of a possible reduction of the subscription rate of *Biometrika* for our members. As a result of these discussions, Professor Pearson offered certain reductions, provided that a sufficient number of subscribers can be secured. Detailed information on this was contained in a memorandum of the

Secretary, P. S. Dwyer, in the November mailing to the membership. It is hoped that many of our members will make use of this opportunity.

With the new constitutions of the American Statistical Association and the Institute of Mathematical Statistics adopted, the way is cleared for the consideration of possible federation plans of the various statistical organizations by the Inter-Society Committee on Federation. J. H. Curtiss and P. S. Olmstead continued to serve as our representatives on the aforementioned committee during 1948. W Feller was our representative on the Policy Committee for Mathematics, and F. C. Mosteller and S. S. Wilks represented the Institute on the Joint Committee for the Development of Statistical Application in Engineering and Manufacturing. W. Bartky was reappointed for a three-year term as our representative to the Division of the Physical Sciences of the National Research Council, and H. Hotelling was our representative to the American Association for the Advancement of Science.

In conclusion, I wish to thank all committee members and others who participated in the work of the Institute during the past year. The heaviest burden falls, of course, on the Secretary and it is hard to express adequately our appreciation for his unselfish efforts and devotion. The smooth and efficient conduct of the affairs of the Institute is largely due to his work.

ABRAHAM WALD
President, 1948

December 31, 1948

REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE FOR 1948[1]

At the beginning of 1948 the Institute had 1037 members and during the period covered by this report 126 new members (13 of whom begin their membership with 1949) joined the Institute and two members were re-instated. During 1948 the Institute lost 64 members of which 24 were by resignation, 38 by suspension for non-payment of dues and 2 by death. Judging from the information available at this date, the Institute will have 1101 members as it starts 1949

Deceased during the year were Dr. Otis A. Pope and H. M. Tompkins.

Meetings of the Institute held during 1948 included those at Columbia University on April 14–15, at the Berkeley campus of the University of California on June 22–24, at the University of Wisconsin on September 6–10, at the University of Washington on November 26–27, and at Cleveland on December 26–30 The Secretary wishes to call attention to the excellent work of the members who served as assistant secretaries at these meetings: Professor Littauer at New York, Professor Arnold at Madison, Professor Birnbaum at Seattle and Professor Van Voorhis at Cleveland.

[1] This report covers the period January 1, 1948 to December 20, 1948 as the books were closed on December 20, 1948 so that the report could be made at the annual meeting.

A summary of the financial transactions of the Institute is given in the *Financial Statement for 1948* which follows:

FINANCIAL STATEMENT

December 31, 1947 to December 20, 1948

A. RECEIPTS

Balance on Hand,[2] December 31, 1947	$5,858.37
Dues	7,482.21
Contributions	255.50
Subscriptions	3,660 40
Sale of Back Numbers	2,718 27
Income from Investments	100 00
Advertising	160.00
Miscellaneous	57 24
Total	$20,291.99

B. EXPENDITURES

Annals—Current		
Office of the Editor	$175.00	
Waverly Press	7,824 66	$7,999.66
Annals—Back Numbers		
Reprinted Vol. XI #2 & #3; XII #2 & #3, XIV #4		1,968 50
Mathematical Reviews and Inter-Society Committee		225 00
Office of the Secretary-Treasurer		
Printing, memoranda, etc. (including some stamped enveloped)	1,174.52	
Postage, supplies, express, telephone calls	225 00	
Clerical help	1,468.00	
Travelling Expense	30.48	2,898.00
Miscellaneous		79.82
Balance on Hand,** December 20, 1948		7,121 01
Total		$20,291.99

C. SUMMARY OF RECEIPTS AND EXPENDITURES

Balance on Hand,** December 31, 1947	$5,858.37
Receipts during 1948	14,433.62
Expenditures during 1948	13,170.98
Balance on Hand,** December 20, 1948	7,121.01

** In bank deposits and government bonds.

D. LIFE MEMBERSHIP FUNDS

It has been the practice to place all life membership payments in a special fund (most of which is in government bonds) and to hold all these funds in reserve until the death of the member—after which his payment is released to the general fund. There were no new life

[2] In bank deposits and government bonds.

membership payments in 1948. During the year a transfer to the general fund has been made of the life membership payment of Professor Irving Fisher, who died in 1947.

	December 31, 1947	*December 20, 1948*
Number of Life Members	30	29
U. S. Government Bonds	$1,888.00	$1,888.00
Bank Deposits	427.00	392.00
Total	$2,315.00	$2,280.00

E BACK ISSUES FUND

It has been our policy, since January 1, 1948, to use income from the sale of back issues to finance the additional reprinting of back issues.

Income from the Sale of Back Issues during 1948	$2,718.27
Expense for Reprinting Back Issues in 1948	1,968.50
Balance in the Fund, December 20, 1948	$749.77

At present 500 copies of Volume 13 #1 and #2 are being reprinted at a cost of $735.00. The payment of this in January will leave a small balance in the fund.

F COMPARISON OF ASSETS ON DECEMBER 31, 1947 AND DECEMBER 20, 1948

	1947	*1948*
U. S. Government G Bonds	$3,000.00	$3,000.00
Life Membership Funds	2,315.00	2,280.00
Back Issues Fund	—	749.77
Additional Bank Deposits	543.47	1,091.24
Current Accounts Receivable	423 55	291.22
Estimated Value (Cost) of Back Annals[3]	10,866.73	12,785.61
Total	$17,148.05	$20,197.84
Net Gain 1948		3,049.19

G LIABILITIES OF INSTITUTE OF MATHEMATICAL STATISTICS AS OF DECEMBER 20, 1948

All bills which have been presented have been paid. The Life Membership Fund now contains $2,280.00 which covers 29 members. Also, $4,060.50 has been paid in for contributions and 1949 dues and subscriptions.

This report does not cover the amount of $13.95 which is held by the Institute for the fund for *Annals* for Countries Devastated by the War (This fund has been under the supervision of Professor Neyman.) During the year this fund purchased $376 25 in back issues (at the agreed rate of $4.50 per volume) which has contributed to the total sales in back issues.

There has been little change in the life membership fund during the year. Our practice of making no transfer of life membership funds until the death of the member is most conservative and protects the interests of the life member.

The question of the value of our inventory is always difficult. We now have 19,083 issues of the *Annals*. At 67¢ per copy, it appears that $12,785.61 is a fair estimate of their actual cost. This is in fact less than 5 times the actual

[3] Cost of *Annals* calculated at 67 cents per copy.

income from back issues this year and hence seems to be a very conservative estimate of the marketable (within ten years) value of our present inventory.

We are in a position now to continue to supply all issues beginning with volume 7 and expect that the sales in back volumes will be such that within two or three years we will be able to reprint the 9 issues in volumes 1–6 which are now practically or completely exhausted.

It appears that the increase in dues and subscriptions has been adequate to take care of the increased expense during 1948. No bonds have been cashed during the year. Additional funds appear necessary for 1949, however, since the present amount of clerical help in the office of the Secretary-Treasurer is utterly inadequate. The employment of additional secretarial assistance, which the Institute must have, will increase the total expense of this office by about $1,200.00. It is necessary, too, to provide a cushion for a possible increase in our Waverly bill, which is up about 10% in 1948. It appears that we may need from $1,500.00 to $2,000.00 additional funds for 1949. Available sources are increases in the number of members and subscribers, contributions from our members, and institutional contributions and memberships.

PAUL S. DWYER
Secretary-Treasurer

December 21, 1948

REPORT OF THE EDITOR FOR 1948

During 1948 the rate of submission of manuscripts for publication in the *Annals* has continued to increase. The size of the *Annals* was held approximately to that set for 1947, the number of pages printed in 1948 being 610. The 1948 volume of the *Annals* contained 59 papers, of which 24 were short notes.

During the past year the backlog of papers has increased to nearly two issues. Thus manuscripts submitted now, especially the longer ones, must wait at least six months after being refereed in order to be printed. If the rate at which manuscripts are submitted increases, as it has during the last two years, this waiting gap may increase to a year by the end of 1949.

If additional funds could be found, it would be highly desirable to increase the *Annals* to 700 pages in 1949.

The manuscripts being received continue to cover a rather wide range of topics in probability and statistics. Almost all of them are research papers. In the Editor's opinion it would be highly desirable for the Institute to take steps, perhaps through invited addresses, to secure good expository and review articles. Sustained attempts have been made over a period of years to obtain such articles by invitation, but with little success.

The Editor wishes to take this opportunity to acknowledge, on behalf of the Editorial Committee, the generous refereeing assistance which has been given by

the following persons: Z W. Birnbaum, A H. Bowker, I W. Burr, G. W. Brown, K. L. Chung, W J Dixon, A. Dvoretzsky, T. N. E Greville, F. E. Grubbs, M. H. Hansen, T. E. Harris, C. Hastings, H. B Horton, G. A. Hunt, B. F. Kimball, T. Koopmans, H. Levene, M. S MacPhail, P. J. McCarthy, R. B. Murphy, M. P. Peisakoff, P. S. Olmstead, E. Paulson, H. G. Romig, L. J. Savage, F. F. Stephan, D F Votaw and J. E. Walsh.

The Editor owes special acknowledgment to Mr. M. E. Freeman for preparation of manuscripts and to Mrs. Frances M. Purvis for other editorial and office assistance.

S. S. WILKS
Editor

December 31, 1948.

STATISTICAL DECISION FUNCTIONS

By Abraham Wald[1]

Columbia University

Introduction and summary. The foundations of a general theory of statistical decision functions, including the classical non-sequential case as well as the sequential case, was discussed by the author in a previous publication [3]. Several assumptions made in [3] appear, however, to be unnecessarily restrictive (see conditions 1–7, pp. 297 in [3]). These assumptions, moreover, are not always fulfilled for statistical problems in their conventional form. In this paper the main results of [3], as well as several new results, are obtained from a considerably weaker set of conditions which are fulfilled for most of the statistical problems treated in the literature. It seemed necessary to abandon most of the methods of proofs used in [3] (particularly those in section 4 of [3]) and to develop the theory from the beginning. To make the present paper self-contained, the basic definitions already given in [3] are briefly restated in section 2.1.

In [3] it is postulated (see Condition 3, p. 207) that the space Ω of all admissible distribution functions F is compact. In problems where the distribution function F is known except for the values of a finite number of parameters, i.e., where Ω is a parametric class of distribution functions, the compactness condition will usually not be fulfilled if no restrictions are imposed on the possible values of the parameters. For example, if Ω is the class of all univariate normal distributions with unit variance, Ω is not compact. It is true that by restricting the parameter space to a bounded and closed subset of the unrestricted space, compactness of Ω will usually be attained. Since such a restriction of the parameter space can frequently be made in applied problems, the condition of compactness may not be too restrictive from the point of view of practical applications. Nevertheless, it seems highly desirable from the theoretical point of view to eliminate or to weaken the condition of compactness of Ω. This is done in the present paper. The compactness condition is completely omitted in the discrete case (Theorems 2.1–2.5), and replaced by the condition of separability of Ω in the continuous case (Theorems 3.1–3.4). The latter condition is fulfilled in most of the conventional statistical problems.

Another restriction postulated in [3] (Condition 4, p. 297) is the continuity of the weight function $W(F, d)$ in F. As explained in section 2.1 of the present paper, the value of $W(F, d)$ is interpreted as the loss suffered when F happens to be the true distribution of the chance variables under consideration and the decision d is made by the statistician. While the assumption of continuity of $W(F, d)$ in F may seem reasonable from the point of view of practical application, it is rather undesirable from the theoretical point of view for the following

[1] Work done under the sponsorship of the Office of Naval Research.

reasons. It is of considerable theoretical interest to consider simplified weight functions $W(F, d)$ which can take only the values 0 and 1 (the value 0 corresponds to a correct decision, and the value 1 to a wrong decision). Frequently, such weight functions are necessarily discontinuous. Consider, for example, the problem of testing the hypothesis H that the mean θ of a normally distributed chance variable X with unit variance is equal to zero. Let d_1 denote the decision to accept H, and d_2 the decision to reject H. Assigning the value zero to the weight W whenever a correct decision is made, and the value 1 whenever a wrong decision is made, we have:

$$W(\theta, d_1) = 0 \text{ for } \theta = 0, \text{ and } = 1 \text{ for } \theta \neq 0; \; W(\theta, d_2) = 0 \text{ for } \theta \neq 0,$$
$$\text{and } = 1 \text{ for } \theta = 0.$$

This weight function is obviously discontinuous. In the present paper the main results (Theorems 2.1–2.5 and Theorems 3.1–3.4) are obtained without making any continuity assumption regarding $W(F, d)$.

The restrictions imposed in the present paper on the cost function of experimentation are considerably weaker than those formulated in [3]. Condition 5 [3, p. 297] concerning the class Ω of admissible distribution functions, and condition 7 [3, p. 298] concerning the class of decision functions at the disposal of the statistician are omitted here altogether.

One of the new results obtained here is the establishment of the existence of so called minimax solutions under rather weak conditions (Theorems 2.3 and 3.2). This result is a simple consequence of two lemmas (Lemmas 2.4 and 3.3) which seem to be of interest in themselves.

The present paper consists of three sections. In the first section several theorems are given concerning zero sum two person games which go somewhat beyond previously published results. The results in section 1 are then applied to statistical decision functions in sections 2 and 3. Section 2 treats the case of discrete chance variables, while section 3 deals with the continuous case. The two cases have been treated separately, since the author was not able to find any simple and convenient way of combining them into a single more general theory.

1. Conditions for strict determinateness of a zero sum two person game. The normalized form of a zero sum two person game may be defined as follows (see [1, section 14.1]): there are two players and there is a bounded and real valued function $K(a, b)$ of two variables a and b given where a may be any point of a space A and b may be any point of a space B. Player 1 chooses a point a in A and player 2 chooses a point b in B, each choice being made in complete ignorance of the other. Player 1 then gets the amount $K(a, b)$ and player 2 the amount $-K(a, b)$. Clearly, player 1 wishes to maximize $K(a, b)$ and player 2 wishes to minimize $K(a, b)$.

Any element a of A will be called a pure strategy of player 1, and any element

b of B a pure strategy of player 2. A mixed strategy of player 1 is defined as follows: instead of choosing a particular element a of A, player 1 chooses a probability measure ξ defined over an additive class $\mathfrak{A}$ of subsets of A and the point a is then selected by a chance mechanism constructed so that for any element α of $\mathfrak{A}$ the probability that the selected element a will be contained in α is equal to $\xi(\alpha)$. Similarly, a mixed strategy of player 2 is given by a probability measure η defined over an additive class $\mathfrak{B}$ of subsets of B and the element b is selected by a chance mechanism so that for any element β of $\mathfrak{B}$ the probability that the selected element b will be contained in β is equal to $\eta(\beta)$. The expected value of the outcome $K(a, b)$ is then given by

$$K^*(\xi, \eta) = \int_B \int_A K(a, b)\, d\xi\, d\eta. \tag{1.1}$$

We can now reinterpret the value of $K(a, b)$ as the value of $K^*(\xi_a, \eta_b)$ where ξ_a and η_b are probability measures which assign probability 1 to a and b, respectively. In what follows, we shall write $K(\xi, \eta)$ for $K^*(\xi, \eta)$, $K(a, b)$ will be used synonymously with $K(\xi_a, \xi_b)$, $K(a, \eta)$ synonymously with $K(\xi_a, \eta)$ and $K(\xi, b)$ synonymously with $K(\xi, \eta_b)$. This can be done without any danger of confusion.

A game is said to be strictly determined if

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta). \tag{1.2}$$

The basic theorem proved by von Neumann [1] states that if A and B are finite the game is always strictly determined, i.e., (1.2) holds. In some previous publications (see [2] and [3]) the author has shown that (1.2) always holds if one of the spaces A and B is finite or compact in the sense of some intrinsic metric, but does not necessarily hold otherwise. A necessary and sufficient condition for the validity of (1.2) was given in [2] for spaces A and B with countably many elements. In this section we shall give sufficient conditions as well as necessary and sufficient conditions for the validity of (1.2) for arbitrary spaces A and B. These results will then be used in later sections.

In what follows, for any subset α of A the symbol ξ_α will denote a probability measure ξ in A for which $\xi(\alpha) = 1$. Similarly, for any subset β of B, ξ_β will stand for a probability measure η in B for which $\eta(\beta) = 1$. We shall now prove the following lemma.

LEMMA 1.1. *Let* $\{\alpha_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *be a sequence of subsets of* A *such that* $\alpha_i \subset \alpha_{i+1}$ *and let* $\alpha = \sum_{i=1}^{\infty} \alpha_i$. *Then*

$$\lim_{i=\infty} \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta). \tag{1.3}$$

PROOF: Clearly, the limit of $\operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta)$ exists as $i \to \infty$ and cannot exceed the value of the right hand member in (1.3). Put

$$\lim_{i=\infty} \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta) = \rho \tag{1.4}$$

and

$$\operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \rho + \delta \qquad (\delta \geq 0). \tag{1.5}$$

Suppose that $\delta > 0$. Then there exists a probability measure ξ_α^0 such that

$$K(\xi_\alpha^0, \eta) \geq \rho + \frac{\delta}{2} \qquad \text{for all } \eta. \tag{1.6}$$

Let $\xi_{\alpha_i}^0$ be the probability measure defined as follows: for any subset α^* of α_i we have

$$\xi_{\alpha_i}^0(\alpha^*) = \frac{\xi_\alpha^0(\alpha^*)}{\xi_\alpha^0(\alpha_i)}. \tag{1.7}$$

Then, since $\lim_{i=\infty} \xi_\alpha^0 (\alpha - \alpha_i) = 0$, we have

$$\lim_{i=\infty} K(\xi_{\alpha_i}^0, \eta) = K(\xi_\alpha^0, \eta) \tag{1.8}$$

uniformly in η. Hence, for sufficiently large i, we have

$$\operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}^0, \eta) \geq \rho + \frac{\delta}{3}, \tag{1.9}$$

which is a contradiction to (1 4). Thus, $\delta = 0$ and Lemma 1.1 is proved. Interchanging the role of the two players, we obtain the following lemma.

LEMMA 1.2. *Let $\{\beta_i\}$ be a sequence of subsets of B such that $\beta_i \subset \beta_{i+1}$ and let $\sum_{i=1}^{\infty} \beta_i = \beta$. Then*

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta_{\beta_i}} \operatorname*{Sup}_{\xi} K(\xi, \eta_{\beta_i}) = \operatorname*{Inf}_{\eta_\beta} \operatorname*{Sup}_{\xi} K(\xi, \eta_\beta). \tag{1.10}$$

We shall now prove the following lemma.

LEMMA 1.3. *The inequality*[2]

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \leq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta) \tag{1.11}$$

always holds.

PROOF: for any given $\epsilon > 0$, it is possible to find probability measures ξ^0 and η^0 such that

$$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta) \geq \operatorname*{Sup}_{\xi} K(\xi, \eta^0) - \epsilon \tag{1.12}$$

and

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \leq \operatorname*{Inf}_{\eta} K(\xi^0, \eta) + \epsilon. \tag{1.13}$$

[2] This inequality was given by v. Neumann [1] for finite spaces A and B.

Then we have,

$$\text{(1.14)}\quad \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta) \leq \operatorname*{Inf}_{\eta} K(\xi^0,\eta) + \epsilon \leq K(\xi^0,\eta^0) + \epsilon$$

$$\leq \operatorname*{Sup}_{\xi} K(\xi,\eta^0) + \epsilon \leqq \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta\quad) + 2\,\epsilon.$$

Since ϵ can be chosen arbitrarily small, Lemma 1.3 is proved.

THEOREM 1.1. *If α is a subset of A such that*

$$\operatorname*{Sup}_{\xi_\alpha}\operatorname*{Inf}_{\eta} K(\xi_\alpha,\eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha,\eta)$$

and

$$\operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha,\eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta),$$

then

$$\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta).$$

PROOF: Clearly,

$$\text{(1.15)}\qquad \operatorname*{Sup}_{\xi_\alpha}\operatorname*{Inf}_{\eta} K(\xi_\alpha,\eta) \leq \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta)$$

and

$$\text{(1.16)}\qquad \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha,\eta) \leq \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta).$$

If the left hand members of (1.15) and (1.16) are equal to each other and equal to the right member of (1.16), then

$$\text{(1.17)}\qquad \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta) \geq \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta).$$

Because of Lemma 1.3 the equality sign must hold and Theorem 1.1 is proved.

Interchanging the two players, we obtain from Theorem 1.1:

THEOREM 1.2. *If β is a subset of B such that* $\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta_\beta} K(\xi,\eta_\beta) = \operatorname*{Inf}_{\eta_\beta}\operatorname*{Sup}_{\xi} K(\xi,\eta_\beta)$ *and* $\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta_\beta} K(\xi,\eta_\beta) = \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta)$,

then

$$\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi,\eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi,\eta).$$

We shall now prove the following theorem.

THEOREM 1.3. *If $\{\alpha_i\}$ is a sequence of subsets of A such that $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$, and if*

$$\text{(1.18)}\qquad \operatorname*{Sup}_{\xi_{\alpha_i}}\operatorname*{Inf}_{\eta} K(\xi_{\alpha_i},\eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i},\eta)$$

for each i, then a necessary and sufficient condition for the validity of

$$\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi, \eta) \tag{1.19}$$

is that

$$\lim_{i=\infty}\operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi} K(\xi, \eta). \tag{1.20}$$

PROOF: Because of (1.18) and Lemma 1.1 we have

$$\lim_{i=\infty}\operatorname*{Inf}_{\eta}\operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi, \eta). \tag{1.21}$$

Hence, (1.20) implies (1.19) and (1.19) implies (1.20). This proves Theorem 1.3.

Interchanging the role of the two players, we obtain from Theorem 1.3 the following theorem.

THEOREM 1.4. *If $\{\beta_i\}$ is a sequence of subsets of B such that $\beta_i \subset \beta_{i+1}$ and $\sum_{i=1}^{\infty} \beta_i = \beta$, and if*

$$\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Inf}_{\eta_{\beta_i}}\operatorname*{Sup}_{\xi} K(\xi, \eta_{\beta_i}),$$

then a necessary and sufficient condition for the validity of (1.19) *is that*

$$\lim_{i=\infty}\operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Sup}_{\xi}\operatorname*{Inf}_{\eta} K(\xi, \eta). \tag{1.22}$$

In [3] an intrinsic metric was introduced in the spaces A and B. The distance of two elements a_1 and a_2 of A is defined by

$$\delta(a_1, a_2) = \operatorname*{Sup}_{b} | K(a_1, b) - K(a_2, b) | . \tag{1.23}$$

Similarly, the distance between two points b_1 and b_2 of B is defined by

$$\delta(b_1, b_2) = \operatorname*{Sup}_{a} | K(a, b_1) - K(a, b_2) | . \tag{1.24}$$

Suppose that there exists a sequence $\{\alpha_i\}$ of subsets of A such that α_i is conditionally compact, $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$.[3] It was shown in [3] that for any conditionally compact subset α_i the relation (1.18) holds. Hence, according to Theorem 1.3, a necessary and sufficient condition for the validity of (1.19) is that (1.20) holds for a sequence $\{\alpha_i\}$ where α_i is conditionally compact, $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$. Similar remarks can be made concerning the space B. The distance definitions given in (1.23) and (1.24) can be extended to the spaces of the probability measures ξ and η, respectively. That is,

$$\delta(\xi_1, \xi_2) = \operatorname*{Sup}_{\eta} | K(\xi_1, \eta) - K(\xi_2, \eta) | \tag{1.25}$$

[3] For a definition of compact and conditionally compact sets, see F. Hausdorff, *Mengenlehre* (3rd edition), p. 107, or [3, p. 296].

and

$$\delta(\eta_1, \eta_2) = \operatorname*{Sup}_{\xi} | K(\xi, \eta_1) - K(\xi, \eta_2) | . \tag{1.26}$$

We shall say that a probability measure ξ is discrete if there exists a denumerable subset α of A such that $\xi(\alpha) = 1$. Similarly, a probability measure η will be said to be discrete if $\eta(\beta) = 1$ for some denumerable subset β of B. We shall now prove the following theorem.

THEOREM 1.5. *If the choice of player* 1 *is restricted to elements of a class* C *of probability measures* ξ *in which the class of all discrete probability measures* ξ *is dense, then a necessary and sufficient condition for the game to be strictly determined is that there exists a sequence* $\{a_i\}$ *of elements of* A *such that*

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta) \tag{1.27}$$

where

$$\alpha_i = \{a_1, a_2, \cdots, a_i\}.$$

PROOF: Since the class of all discrete probability measures ξ lies dense in the class C, there exists a sequence $\alpha = \{a_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ such that

$$\operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta). \tag{1.28}$$

Since $\alpha_i = \{a_1, \cdots, a_i\}$ is finite, we have

$$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta). \tag{1.29}$$

It then follows from Lemma 1.1 that

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta). \tag{1.30}$$

Clearly, (1.30) and strict determinateness of the game implies (1.27). On the other hand, any $\alpha = \{a_i\}$ that satisfies (1.27), will satisfy also (1.28) and (1.30). But (1.27) and (1.30) imply that the game is strictly determined. Thus, Theorem 1.5 is proved.

THEOREM 1.6. *If the choice of player* 2 *is restricted to elements of a class* C *of probability measure* η *in which the class of all discrete probability measures* η *lies dense, then a necessary and sufficient condition for the strict determinateness of the game is that there exists a sequence* $\beta = \{b_i\}$ *of elements of* B *such that*

$$\lim_{i=\infty} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \tag{1.31}$$

where

$$\beta_i = \{b_1, \cdots, b_i\}.$$

This theorem is obtained from Theorem 1.5 by interchanging the players 1 and 2.

2. Statistical decision functions: the case of discrete chance variable.

2.1. *The problem of statistical decisions and its interpretation as a zero sum two person game.* In some previous publications (see, for example, [3]) the author has formulated the problem of statistical decisions as follows: Let $X = \{X^i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be an infinite sequence of chance variables. Any particular observation x on X is given by a sequence $x = \{x^i\}$ of real values where x^i denotes the observed value of X^i. Suppose that the probability distribution $F(x)$ of X is not known. It is, however, known that F is an element of a given class Ω of distribution functions. There is, furthermore, a space D given whose elements d represent the possible decisions that can be made in the problem under consideration. Usually each element d of D will be associated with a certain subset ω of Ω and making the decision d can be interpreted as accepting the hypothesis that the true distribution is included in the subset ω. The fundamental problem in statistics is to give a rule for making a decision, that is, a rule for selecting a particular element d of D on the basis of the observed sample point x. In other words, the problem is to construct a function $d(x)$, called decision function, which associates with each sample point x an element $d(x)$ of D so that the decision $d(x)$ is made when the sample point x is observed.

This formulation of the problem includes the sequential as well as the classical non-sequential case. For any sample point x, let $n(x)$ be the number of components of x that must be known to be able to determine the value of $d(x)$. In other words, $n(x)$ is the smallest positive integer such that $d(y) = d(x)$ for any y whose first n coordinates are equal to the first n coordinates of x. If no finite n exists with the above property, we put $n = \infty$. Clearly, $n(x)$ is the number of observations needed to reach a decision. To put in evidence the dependence of $n(x)$ on the decision rule used, we shall occasionally write $n(x; \mathfrak{D})$ instead of $n(x)$ where $\mathfrak{D}$ denotes the decision function $d(x)$ used. If $n(x)$ is constant over the whole sample space, we have the classical case, that is the case where a decision is to be made on the basis of a predetermined number of observations. If $n(x)$ is not constant over the sample space, we have the sequential case. A basic question in statistics is this: What decision function should be chosen by the statistician in any given problem? To set up principles for a proper choice of a decision function, it is necessary to express in some way the degree of importance of the various wrong decisions that can be made in the problem under consideration. This may be expressed by a non-negative function $W(F, d)$, called weight functions, which is defined for all elements F of Ω and all elements d of D. For any pair (F, d), the value $W(F, d)$ expresses the loss caused by making the decision d when F is the true distribution of X. For any positive integer n, let $c(n)$ denote the cost of making n observations. If the decision function $\mathfrak{D} = d(x)$ is used the expected loss plus the expected cost of experimentation is given by

$$r[F, \mathfrak{D}] = \int_M W[F, d(x)]\, dF(x) + \int_M c(n(x))\, dF(x) \tag{2.1}$$

where M denotes the sample space, i.e. the totality of all sample points x. We shall use the symbol $\mathfrak{D}$ for $d(x)$ when we want to indicate that we mean the whole decision function and not merely a value of $d(x)$ coresponding to some x.

The above expression (2.1) is called the risk. Thus, the risk is a real valued non-negative function of two variables F and $\mathfrak{D}$ where F may be any element of Ω and $\mathfrak{D}$ any decision rule that may be adopted by the statistician.

Of course, the statistician would like to make the risk r as small as possible. The difficulty he faces in this connection is that r depends on two arguments F and $\mathfrak{D}$, and he can merely choose $\mathfrak{D}$ but not F. The true distribution F is chosen, we may say, by Nature and Nature's choice is usually entirely unknown to the statistician. Thus, the situation that arises here is very similar to that of a zero sum two person game. As a matter of fact, the statistical problem may be interpreted as a zero sum two person game by setting up the following correspondence.

Two Person Game	*Statistical Decision Problem*
Player 1	Nature
Player 2	Statistician
Pure strategy a of player 1	Choice of true distribution F by Nature
Pure strategy b of player 2	Choice of decision rule $\mathfrak{D} = d(x)$
Space A	Space Ω
Space B	Space Q of decision rules $\mathfrak{D}$ that can be used by the statistician.
Outcome $K(a, b)$	Risk $r(F, \mathfrak{D})$
Mixed strategy ξ of player 1	Probability measure ξ defined over an additive class of subsets of Ω (a priori probability distribution in the space Ω)
Mixed strategy η of player 2	Probability measure η defined over an additive class of subsets of the space Q. We shall refer to η as randomized decision function.
Outcome $K(\xi, \eta)$ when mixed strategies are used.	Risk $r(\xi, \eta) = \int_Q \int_\Omega r(F, \mathfrak{D})\, d\xi\, d\eta$.

2.2. *Formulation of some conditions concerning the spaces Ω, D, the weight function $W(F, d)$ and the cost function of experimentation.* A general theory of statistical decision functions was developed in [3] assuming the fulfillment of seven conditions listed on pp. 297–8.[4] The conditions listed there are unnecessarily restrictive and we shall replace them here by a considerably weaker set of conditions.

In this chapter we shall restrict ourselves to the study of the case where each of the chance variables $X^1, X^2, \cdots$, ad inf. is discrete. We shall say that a chance

[4] In [3] only the continuous case is treated (existence of a density function is assumed), but all the results obtained there can be extended without difficulty to the discrete case.

variable is discrete if it can take only countably many different values. Let $a_{i1}, a_{i2}, \cdots$, ad inf. denote the possible values of the chance variable X^i. Since it is immaterial how the values a_{ij} are labeled, there is no loss of generality in putting $a_{ij} = j (j = 1, 2, 3, \cdots$, ad inf.). Thus, we formulate the following condition.

CONDITION 2.1. *The chance variable X^i ($i = 1, 2, \cdots$, ad inf.) can take only positive integral values.*

As in [3], also here we postulate the boundedness of the weight function, i.e., we formulate the following condition.

CONDITION 2.2. *The weight function $W(F, d)$ is a bounded function of F and d.*

To formulate condition 2.3, we shall introduce some definitions. Let ω be a given subset of Ω. The distance between two elements d_1 and d_2 of D relative to ω is defined by

$$\delta(d_1, d_2 ; \omega) = \sup_{F \in \omega} | W(F, d_1) - W(F, d_2) | . \tag{2.2}$$

We shall refer to $\delta(d_1, d_2 ; \Omega)$ as the absolute distance, or more briefly, the distance between d_1 and d_2. We shall say that a subset D^* of D is compact (conditionally compact) relative to ω, if it is compact (conditionally compact) in the sense of the metric $\delta(d_1, d_2 ; \omega)$. If D^* is compact relative to Ω, we shall say briefly that D^* is compact.

An element d of D is said to be uniformly better than the element d' of D relative to a subset ω of Ω if

$$W(F, d) \leqq W(F, d') \text{ for all } F \text{ in } \omega$$

and if

$$W(F, d) < W(F, d') \text{ for at least one } F \text{ in } \omega.$$

A subset D^* of D is said to be complete relative to a subset ω of Ω if for any d outside D^* there exists an element d^* in D^* such that d^* is uniformly better than d relative to ω.

CONDITION 2.3. *For any positive integer i and for any positive ϵ there exists a subset $D^*_{i,\epsilon}$ of D which is compact relative to Ω and complete relative to $\omega_{i,\epsilon}$ where $\omega_{i,\epsilon}$ is the class of all elements F of Ω for which prob $\{X^1 \leqq i\} \geqq \epsilon$.*

If D is compact, then it is compact with respect to any subset ω of Ω and Condition 2.3 is fulfilled. For any finite space D, Condition 2.3 is obviously fulfilled. Thus, Condition 2.3 is fulfilled, for example, for any problem of testing a statistical hypothesis H, since in that case the space D contains only two elements d_1 and d_2 where d_1 denotes the decision to reject H and d_2 the decision to accept H.

In [3] it was assumed that the cost of experimentation depends only on the number of observations made. This assumption is unnecessarily restrictive. The cost may depend also on the decision rule $\mathfrak{D}$ used. For example, let $\mathfrak{D}_1$ and $\mathfrak{D}_2$ be two decision rules such that $n(x; \mathfrak{D}_1)$ is equal to a constant n_0, while

$\mathfrak{D}_2$ is such that at any stage of the experimentation where $\mathfrak{D}_2$ requires taking at least one additional observation the probability is positive that experimentation will be terminated by taking only one more observation. Let x^0 be a particular sample point for which $n(x^0; \mathfrak{D}_2) = n(x^0, \mathfrak{D}_1) = n_0$. There are undoubtedly cases where the cost of experimentation is appreciably increased by the necessity of having to look at the observations at each stage of the experiment before we can decide whether or not to continue taking additional observations Thus in many cases the cost of experimentation when x^0 is observed may be greater for $\mathfrak{D}_2$ than for $\mathfrak{D}_1$. The cost may also depend on the actual values of the observations made. Thus, we shall assume that the cost c is a single valued function of the observations $x^1, \cdots, x^m$ and the decision rule $\mathfrak{D}$ used, i.e., $c = c(x^1, \cdots, x^m, \mathfrak{D})$.

CONDITION 2.4. *The cost* $c(x^1, \cdots, x^m, \mathfrak{D})$ *is non-negative and* $\lim c(x^1, \cdots, x^m, \mathfrak{D}) = \infty$ *uniformly in* $x^1, \cdots, x^m, \mathfrak{D}$ *as* $m \to \infty$. *For each positive integral value* m, *there exists a finite value* c_m, *depending only on* m, *such that* $c(x^1, \cdots, x^m, \mathfrak{D}) \leqq c_m$ *identically in* $x^1, \cdots, x^m, \mathfrak{D}$. *Furthermore,* $c(x^1, \cdots, x^m, \mathfrak{D}_1) = c(x^1, \cdots, x^m, \mathfrak{D}_2)$ *if* $n(x; \mathfrak{D}_1) = n(x; \mathfrak{D}_2)$ *for all* x. *Finally, for any sample point* x *we have* $c(x^1, \cdots, x^{n(x,\mathfrak{D}_1)}, \mathfrak{D}_1) \leqq c(x^1, \cdots, x^{n(x,\mathfrak{D}_2)}, \mathfrak{D}_2)$ *if there exists a positive integer* m *such that* $n(x, \mathfrak{D}_1) = n(x, \mathfrak{D}_2)$ *when* $n(x, \mathfrak{D}_2) < m$ *and* $n(x, \mathfrak{D}_1) = m$ *when* $n(x, \mathfrak{D}_2) \geqq m$.

2.3 *Alternative definition of a randomized decision function, and a further condition on the cost function.* In Section 2.1 we defined a randomized decision function as a probability measure η defined over some additive class of subsets of the space Q of all decision functions $d(x)$. Before formulating an alternative definition of a randomized decision function, we have to make precise the meaning of η by stating the additive class C_Q of subsets of Q over which η is defined. Let C_D be the smallest additive class of subsets of D which contains all subsets of D which are open in the sense of the metric $\delta(d_1, d_2; \Omega)$. For any finite set of positive integers $a_1, \cdots, a_k$ and for any element D^* of C_D, let $Q(a_1, \cdots, a_k, D^*)$ be the set of all decision functions $d(x)$ which satisfy the following two conditions: (1) If $x^1 = a_1, x^2 = a_2, \cdots, x^k = a_k$, then $n(x) = k$; (2) If $x^1 = a_1, \cdots, x^k = a_k$, then $d(x)$ is an element of D^*. Let C_Q^* be the class of all sets $Q(a_1, \cdots, a_k, D^*)$ corresponding to all possible values of k, $a_1, \cdots, a_k$ and all possible elements D^* of C_D. The additive class C_Q is defined as the smallest additive class containing C_Q^* as a subclass. Then with any η we can associate two sequences of functions

$$\{z_m(x^1, \cdots, x^m \mid \eta)\}$$

and

$$\{\delta_{x^1, \cdots, x^m}(D^* \mid \eta)\} (m = 1, 2, \cdots, \text{ad inf.})$$

where $0 \leqq z_m(x^1, \cdots, x^m \mid \eta) \leqq 1$ and for any $x^1, \cdots, x^m$, $\delta_{x^1 \cdots x^m}$ is a probability measure in D defined over the additive class C_D. Here

$$z_m(x^1, \cdots, x^m \mid \eta)$$

denotes the conditional probability that $n(x) > m$ under the condition that the first m observations are equal to $x^1, \cdots, x^m$ and experimentation has not been terminated for $(x^1, \cdots, x^k)$ for $(k = 1, 2, \cdots, m-1)$, while

$$\delta_{x^1,\cdots,x^m}(D^* \mid \eta)$$

is the conditional probability that the final decision d will be an element of D^* under the condition that the sample $(x^1, \cdots, x^m)$ is observed and $n(x) = m$. Thus

$$z_1(x^1 \mid \eta) z_2(x^1, x^2 \mid \eta) \cdots z_{m-1}(x^1, \cdots, x^{m-1} \mid \eta)\,[1 - z_m(x^1, \cdots, x^m \mid \eta)] =$$

(2.3)

$$\eta[Q(x^1, \cdots, x^m, D]$$

and

$$\delta_{x^1 \cdots x^m}(D^* \mid \eta) = \frac{\eta[Q(x^1, \cdots, x^m, D^*)]}{\eta[Q(x^1, \cdots, x^m, D)]}. \tag{2.4}$$

We shall now consider two sequences of functions $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$, not necessarily generated by a given η. An alternative definition of a randomized decision function can be given in terms of these two sequences as follows: After the first observation x^1 has been drawn, the statistician determines whether or not experimentation be continued by a chance mechanism constructed so that the probability of continuing experimentation is equal to $z_1(x^1)$. If it is decided to terminate experimentation, the statistician uses a chance mechanism to select the final decision d constructed so that the probability distribution of the selected d is equal to $\delta_{x^1}(D^*)$. If it is decided to take a second observation and the value x^2 is obtained, again a chance mechanism is used to determine whether or not to stop experimentation such that the probability of taking a third observation is equal to $z_2(x^1, x^2)$. If it is decided to stop experimentation, a chance mechanism is used to select the final d so that the probability distribution of the selected d is equal to $\delta_{x^1 x^2}(D^*)$, and so on.

We shall denote by ζ a randomized decision function defined in terms of two sequences $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$, as described above. Clearly, any given η generates a particular ζ. Let $\zeta(\eta)$ denote the ζ generated by η. One can easily verify that two different η's may generate the same ζ, i.e., there exist two different η's, say η_1 and η_2 such that $\zeta(\eta_1) = \zeta(\eta_2)$.

We shall now show that for any ζ there exists an η such that $\zeta(\eta) = \zeta$. Let ζ be given by the two sequences $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$. Let b_j denote a sequence of r_j positive integers, i.e., $b_j = (b_{j1}, \cdots, b_{jr_j})$ $(j = 1, 2, \cdots, k)$ subject to the restriction that no b_j is equal to an initial segment of $b_l (j \neq l)$. Let, furthermore, $D_1^*, \cdots, D_k^*$ be k elements of C_D. Finally, let $Q(b_1, \cdots, b_k, D_1^*, \cdots, D_k^*)$ denote the class of all decision functions $d(x)$ which satisfy

the following condition: If $(x^1, \cdots, x^{r_j}) = b_j$ then $n(x) = r_j$ and $d(x)$ is an element of $D_j^*(j = 1, \cdots, k)$. Let η be a probability measure such that

$$\eta[Q(b_1, \cdots, b_k, D_1^*, \cdots, D_k^*)]$$

$$= \delta_{b_1}(D_1^*) \cdots \delta_{b_k}(D_k^*) \prod_{m=1}^{\infty} \prod_{x^m=1}^{\infty} \prod_{x^{m-1}=1}^{\infty} \cdots \prod_{x^1=1}^{\infty} \cdot \{z_m(x^1, \cdots, x^m)^{g_m(x^1,\cdots,x^m)}[1 - z_m(x^1, \cdots, x^m)]^{g^*_m(x^1,\cdots,x^m)}\} \tag{2.5}$$

holds for all values of $k, b_1, \cdots, b_k, D_1^*, \cdots, D_k^*$. Here $g_m(x^1, \cdots, x^m) = 1$ if $(x^1, \cdots, x^m)$ is equal to an initial segment of at least of one of the samples $b_1, \cdots, b_k$, but is not equal to any of the samples $b_1, \cdots, b_k$. In all other cases $g_m(x^1, \cdots, x^m) = 0$. The function $g^*_m(x^1, \cdots, x^m)$ is equal to 1 if $(x^1, \cdots, x^m)$ is equal to one of the samples $b_1, \cdots, b_k$, and zero otherwise. Clearly, for any η which satisfies (2.5) we have $\zeta(\eta) = \zeta$. The existence of such an η can be shown as follows. With any finite set of positive integers $i_1, \cdots, i_r$ we associate an elementary event, say $A_r(i_1, \cdots, i_r)$. Let $\bar{A}_r(i_1, \cdots, i_r)$ denote the negation of the event $A_r(i_1, \cdots, i_r)$. Thus, we have a denumerable system of elementary events by letting $r, i_1, \cdots, i_r$ take any positive integral values. We shall assume that the events $A_1(1), A_1(2), \cdots$, ad inf. are independent and the probability that $A_1(i)$ happens is equal $z_1(i)$. We shall now define the conditional probability of $A_2(i, j)$ knowing for any k whether $A_1(k)$ or $\bar{A}_1(k)$ happened. If $A_1(i)$ happened, the conditional probability of $A_2(i, j) = z_2(i, j)$ and 0 otherwise. The conditional probability of the joint event that $A_2(i_1, j_1), A_2(i_2, j_2), \cdots, A_2(i_r, j_r), \bar{A}_2(i_{r+1}, j_{r+1}), \cdots$, and $\bar{A}_2(i_{r+s}, j_{r+s})$ will happen is the product of the conditional probabilities of each of these events (knowing for each i whether $A_1(i)$ or $\bar{A}_1(i)$ happened). Similarly, the conditional probability (knowing for any i and for any (i, j), whether the corresponding event $A_2(i, j)$ happened or not) that $A_3(i_1, j_1, k_1)$ and $A_3(i_2, j_2, k_2)$ and $\cdots A_3(i_r, j_r, k_r)$ and $\bar{A}_3(i_{r+1}, j_{r+1}, k_{r+1})$ and $\cdots$ and $\bar{A}_3(i_{r+s}; j_{r+s}, k_{r+s})$ will simultaneously happen is equal to the product of the conditional probabilities of each of them. The conditional probability of $A_3(i, j, k)$ is equal to $z_3(i, j, k)$ if $A_1(i)$ and $A_2(i, j)$ happened, and zero otherwise; and so on. Clearly, this system of probabilities is consistent.

If we interpret $A_r(i_1, \cdots, i_r)$ as the event that the decision function $\mathfrak{D} = d(x)$ selected by the statistician has the property that $n(x; \mathfrak{D}) > r$ when $x^1 = i_1, \cdots, x^r = i_r$, the above defined system of probabilities for the denumerable sequence $\{A_r(i_1, \cdots, i_r)\}$ of events implies the validity of (2.5) for $D_j^* = D(j = 1, \cdots, k)$. The consistency of the formula (2.5) for $D_j^* = D$ implies, as can easily be verified, the consistency of (2.5) also in the general case when $D_j^* \neq D$.

Let ζ_i be given by the sequences of $\{z_{mi}(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m, i}\}$ $(m = 1, 2, \cdots$, ad inf.). Let, furthermore, ζ be given by $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}\}$. We shall say that

$$\lim_{i=\infty} \zeta_i = \zeta \tag{2.6}$$

if for any m, $x^1, \cdots, x^m$ we have

$$\lim_{i=\infty} z_{m,i}(x^1, \cdots, x^m) = z_m(x^1, \cdots, x^m) \tag{2.7}$$

and

$$\lim_{i=\infty} \delta_{x^1 \cdots x^m, i}(D^*) = \delta_{x^1 \cdots x^m}(D^*) \tag{2.8}$$

for any open subset D^* of D whose boundary has probability measure zero according to the limit probability measure $\delta_{x^1 \cdots x^m}$.

In addition to Condition 2.4, we shall impose the following continuity condition on the cost function.

CONDITION 2.5. *If*

$$\lim_{i=\infty} \zeta(\eta_i) = \zeta(\eta),$$

then

$$\lim_{i=\infty} \int_{Q(x^1, \cdots, x^m)} c(x^1, \cdots, x^m, \mathfrak{D})\, d\eta_i = \int_{Q(x^1, \cdots, x^m)} c(x^1, \cdots, x^m, \mathfrak{D})\, d\eta.$$

where $Q(x^1, \cdots, x^m)$ is the class of all decision functions $\mathfrak{D}$ for which $n(y, \mathfrak{D}) = m$ if $y^1 = x^1, \cdots, y^m = x^m$.

2.4. *The main theorem.* In this section we shall show that the statistical decision problem, viewed as a zero sum two person game, is strictly determined. It will be shown in subsequent sections that this basic theorem has many important consequences for the theory of statistical decision functions. A precise formulation of the theorem is as follows:

THEOREM 2.1. *If Conditions 2.1–2.5 are fulfilled, the decision problem, viewed as a zero sum two person game, is strictly determined, i.e.,*

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r(\xi, \eta). \tag{2.9}$$

To prove the above theorem, we shall first derive several lemmas.

LEMMA 2.1. *For any $\epsilon > 0$, there exists a positive integer m_ϵ, depending only on ϵ, such that the value of $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta)$, is not changed by more than ϵ if we restrict the choice of the statistician to decision functions $d(x)$ for which $n(x) \leqq m_\epsilon$ for all x.*

PROOF: Put $W_0 = \operatorname*{Sup}_{F, D} W(F, d)$ and choose m_ϵ so that

$$c(x^1, \cdots, x^m, \mathfrak{D}) > \frac{W_0^2}{\epsilon} \tag{2.10}$$

identically in $x^1, \cdots, x^m$ and $\mathfrak{D}$ for all $m \geqq m_\epsilon$. The existence of such a value m_ϵ follows from Condition 2.4. Consider the function $\operatorname*{Inf}_{\mathfrak{D}} r(\xi, \mathfrak{D})$. Our lemma is proved, if we can show that for any ξ, the value of $\operatorname*{Inf}_{\mathfrak{D}} r(\xi, \mathfrak{D})$ is not increased

by more than ϵ if we restrict $\mathfrak{D}$ to be such that $n(x, \mathfrak{D}) \leq m_\epsilon$ for all x. The latter statement is proved, if we can show that for any decision function $\mathfrak{D}_1 = d_1(x)$ we can find another decision function $\mathfrak{D}_2 = d_2(x)$ such that $n(x, \mathfrak{D}_2) \leq m_\epsilon$ for all x and $r(\xi, \mathfrak{D}_2) \leq r(\xi, \mathfrak{D}_1) + \epsilon$. There are two cases to be considered: (a) prob $\{n(X, \mathfrak{D}_1) > m_\epsilon \mid \xi\} \geqq \epsilon/W_0$ and (b) prob $\{n(X, \mathfrak{D}_1) > m_\epsilon \mid \xi\} < \epsilon/W_0$. In case (a) we have $r(\xi, \mathfrak{D}_1) \geqq W_0$. In this case we can choose $\mathfrak{D}_2$ to be the rule that we decide for some element d_0 of D without taking any observations. Clearly, for this choice of $\mathfrak{D}_2$ we shall have $r(\xi, \mathfrak{D}_2) \leq r(\xi, \mathfrak{D}_1)$. In case (b) we choose $\mathfrak{D}_2$ as follows:

$$d_2(x) = d_1(x) \text{ whenever } n(x, \mathfrak{D}_1) \leqq m_\epsilon ;$$

$$d_2(x) = d_0 \text{ whenever } n(x, \mathfrak{D}_1) > m_\epsilon ,$$

where d_0 is an arbitrary element of D. Thus, $n(x, \mathfrak{D}_2) \leqq m_\epsilon$ for all x. Since prob $\{n(x, \mathfrak{D}_1) > m_\epsilon \mid \xi) < \epsilon/W_0$, it is clear that $r(\xi, \mathfrak{D}_2) \leqq r(\xi, \mathfrak{D}_1) + \epsilon$. Hence our lemma is proved.

Let Q^m denote the class of decision functions $\mathfrak{D}$ for which $n(x; \mathfrak{D}) \leqq m$ for all x. For any positive ϵ, let $Q^{m,\epsilon}$ denote the class of all decision functions which satisfy the following two conditions simultaneously: (1) $n(x, \mathfrak{D}) \leqq m$ for all x; (2) $d(x)$ is an element of $D^*_{x^1,\epsilon}$ where $D^*_{x^1,\epsilon}$ denotes the subset of D having the properties stated in Condition 2.3. Clearly, $Q^{m,\epsilon} \subset Q^m$. A probability measure η will be denoted by η^m if $\eta(Q^m) = 1$, and by $\eta^{m,\epsilon}$ if $\eta(Q^{m,\epsilon}) = 1$.

LEMMA 2.2. *The following inequality holds:*

$$\text{(2.11)} \quad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r(\xi, \eta^m) \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,\epsilon}} r(\xi, \eta^{m,\epsilon}) \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r(\xi, \eta^m) + \epsilon W_0 ,$$

where W_0 is an upper bound of $W(F, d)$.

PROOF: The first half of (2.11) is obvious. If we replace the subscript x^1 by the chance variable X^1, the set $\omega_{x^1,\epsilon}$ defined in Condition 2.3 will be a random subset of Ω. It follows easily from the definition of $\omega_{x^1,\epsilon}$ that

$$\text{(2.12)} \qquad \text{prob } \{F \epsilon \omega_{X^1,\epsilon} \mid F\} \geqq 1 - \epsilon.$$

With any decision function $\mathfrak{D} = d(x)$ we shall associate another decision function $\mathfrak{D}^* = d^*(x)$ such that $n(x, \mathfrak{D}) = n(x, \mathfrak{D}^*)$; $d^*(x) = d(x)$ whenever $d(x) \epsilon D^*_{x^1,\epsilon}$; and $d^*(x)$ is an element of $D^*_{x^1,\epsilon}$ that is uniformly better than $d(x)$ relative to $\omega_{x^1,\epsilon}$ whenever $d(x) \notin D^*_{x^1,\epsilon}$. It follows from (2.12) and the fact that W_0 is an upper bound of $W(F, d)$ that

$$\text{(2.13)} \qquad r(F, \mathfrak{D}^*) \leqq r(F, \mathfrak{D}) + \epsilon W_0.$$

The second half of (2.11) is an immediate consequence of (2.13) and our lemma is proved.

LEMMA 2.3. *The equation*

$$\text{(2.14)} \qquad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,\epsilon}} r(\xi, \eta^{m,\epsilon}) = \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi} r(\xi, \eta^{m,\epsilon})$$

holds for all m and ϵ.

PROOF. For any positive integral values m, k and for any $\rho > 0$, let $\Omega^{m,k,\rho}$ be the class of all elements F of Ω for which

$$\text{prob } \{x^1 \leqq k \text{ and } x^2 \leqq k \text{ and } \cdots x^m \leqq k\} \geqq 1 - \rho.$$

A probability measure ξ for which $\xi(\Omega^{m,k,\rho}) = 1$ will be denoted by $\xi^{m,k,\rho}$. To prove (2.14), we shall first prove the inequality

$$(2.15)\quad \left| \operatorname*{Sup}_{\xi^{m,k,\rho}} \operatorname*{Inf}_{\eta^{m,\epsilon}} r(\xi^{m,k,\rho}, \eta^{m,\epsilon}) - \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta^{m,\epsilon}) \right| \leqq \rho(W_0 + C_m)$$

where C_m is an upper bound of $C(x^1, \cdots, x^r, \mathfrak{D})$ for all $r \leqq m$, $x^1, \cdots, x^r$ and $\mathfrak{D}$.

Since for any $d(x)$ in $Q^{m,\epsilon}$, $d(x)$ must be an element of $D^*_{x^1,\epsilon}$ and since $D^*_{x^1,\epsilon}$ is compact, it is sufficient to prove the validity of (2.14) in the case when $D^*_{x^1,\epsilon}$ is a finite set Thus, we shall assume in the remainder of the proof that $D^*_{x^1,\epsilon}$ is finite

Let δ be a given positive number and let $Q^{m,k,\epsilon}$ be a finite subset of $Q^{m,\epsilon}$ satisfying the following condition: for any element $\mathfrak{D} = d(x)$ in $Q^{m,\epsilon}$ there exists an element $\mathfrak{D}^* = d^*(x)$ in $Q^{m,k,\epsilon}$ such that

$$d^*(x) = d(x) \qquad \text{and} \qquad |C(x, \mathfrak{D}^*) - C(x, \mathfrak{D})| \leqq \delta$$

for all x for which $x^1 \leqq k$, $x^2 \leqq k$, $\cdots$, and $x^m \leqq k$. Clearly, for any choice of δ there exists a finite subset $Q^{m,k,\epsilon}$ of $Q^{m,\epsilon}$ with the desired property. For any $\mathfrak{D}$ in $Q^{m,\epsilon}$, we can then find an element $\mathfrak{D}^*$ in $Q^{m,k,\epsilon}$ such that

$$r(F, \mathfrak{D}^*) \leqq r(F, \mathfrak{D}) + \rho(W_0 + C_m) + \delta,$$

for all F in $\Omega^{m,k,\rho}$. From this it follows that

$$(2.16)\quad \operatorname*{Sup}_{\xi^{m,k,\rho}} \operatorname*{Inf}_{\eta^{m,\epsilon}} r \leqq \operatorname*{Sup}_{\xi^{m,k,\rho}} \operatorname*{Inf}_{\eta^{m,k,\epsilon}} r \leqq \operatorname*{Sup}_{\xi^{m,k,\rho}} \operatorname*{Inf}_{\eta^{m,\epsilon}} r + \rho(W_0 + C_m) + \delta$$

$$(2.17)\quad \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r \leqq \operatorname*{Inf}_{\eta^{m,k,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r \leqq \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r + \rho(W_0 + C_m) + \delta$$

where $\eta^{m,k,\epsilon}(Q^{m,k,\epsilon}) = 1$. Since $Q^{m,k,\epsilon}$ is finite, we have

$$(2.18)\qquad \operatorname*{Sup}_{\xi^{m,k,\rho}} \operatorname*{Inf}_{\eta^{m,k,\epsilon}} r = \operatorname*{Inf}_{\eta^{m,k,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r.$$

Inequality (2.15) follows from (2.16), (2.17) and (2.18) and the fact that δ can be chosen arbitrarily small.

Lemmas 1.1., 1.3 and the inequality (2.15) imply that Lemma 2.3 must hold if

$$(2.19)\qquad \lim_{k=\infty} \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r = \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi} r$$

holds. Thus, the proof of Lemma 2.3 is completed if we can show the validity of (2.19).

Let $\{\eta_k^{m,\epsilon}\}$ ($k = 1, 2, \cdots$, ad inf.) be a sequence of randomized decision functions such that

$$(2.20)\qquad \lim_{k=\infty} \left[\operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta_k^{m,\epsilon}) - \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta^{m,\epsilon})\right] = 0.$$

Let $\zeta_k = \zeta(\eta_k^{m,e})$ (see definition in Section 3.2) and let ζ_k be given by the two sequences of functions $\{z_{rk}(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r, k}\}$ $(r = 1, 2, \cdots, m)$. Since there are only countably many samples $(x^1, \cdots, x^r)$ $(r \leqq m)$, there exists a subsequence $\{k^1\}$ of the sequence $\{k\}$ such that

$$\lim_{k=\infty} z_{r,k^1}(x^1, \cdots, x^r) = z_r(x^1, \cdots, x^r) \tag{2.21}$$

and

$$\lim_{k=\infty} \delta_{x^1 \cdots x^r, k^1} = \delta_{x^1 x^2 \cdots x^r} \tag{2.22}$$

for all r and all samples $(x^1, \cdots, x^r)$. Let $\eta_0^{m,e}$ be a randomized decision function such that $\zeta(\eta_0^{m,e})$ is equal to the ζ defined by $\{z_r(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r}\}$ $(r = 1, 2, \cdots, m)$.

For any element F of Ω and for any $\nu > 0$, there exists a finite subset M_ν of the m-dimensional sample space such that the probability (under F) that the sample $(x^1, \cdots, x^m)$ will fall in M_ν is $\geqq 1 - \nu$. From this and the continuity of the cost function (Condition 2.5) it follows that

$$\lim_{k=\infty} r(F, \eta_{k^1}^{m,e}) = r(F, \eta_0^{m,e}) \text{ for all } F. \tag{2.23}$$

Clearly,

$$\operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta) = \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta) \tag{2.24}$$

where $F^{m,k,\rho}$ is an element of $\Omega^{m,k,\rho}$. Hence

$$\operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta^{m,e}) = \operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,e}). \tag{2.25}$$

Since any F in Ω is contained in $\Omega^{m,k,\rho}$ for sufficiently large k, it follows from (2.20) and (2.25) that

$$\lim_{k=\infty} r(F, \eta_{k^1}^{m,e}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,e})\}. \tag{2.26}$$

Hence, because of (2.23),

$$r(F, \eta_0^{m,e}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,e})\}. \tag{2.27}$$

Thus,

$$\operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{F} r(F, \eta^{m,e}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,e}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,e})\}. \tag{2.28}$$

Since the left hand member of (2.28) cannot be smaller than the right hand member, the equality sigh must hold. This concludes the proof of Lemma 2.3.

Theorem 2.1 can easily be proved with the help of lemmas 2.1, 2.2 and 2.3. From Lemma 2.2 it follows that

$$\lim_{e=0} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,e}} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r. \tag{2.29}$$

From this and Lemma 2.3 we obtain

$$\lim_{\epsilon=0} \operatorname*{Inf}_{\eta^{m},\epsilon} \operatorname*{Sup}_{\xi} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m}} r. \tag{2.30}$$

But

$$\lim_{\epsilon=0} \operatorname*{Inf}_{\eta^{m},\epsilon} \operatorname*{Sup}_{\xi} r \geqq \operatorname*{Inf}_{\eta^{m}} \operatorname*{Sup}_{\xi} r.$$

Hence

$$\operatorname*{Inf}_{\eta^{m}} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m}} r. \tag{2.31}$$

Hence, because of Lemma 1.3, we then must have

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m}} r = \operatorname*{Inf}_{\eta^{m}} \operatorname*{Sup}_{\xi} r. \tag{2.32}$$

It follows from Lemma 2.1 that

$$\lim_{m=\infty} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m}} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r. \tag{2.33}$$

Hence, because of (2.32), we have

$$\lim_{m=\infty} \operatorname*{Inf}_{\eta^{m}} \operatorname*{Sup}_{\xi} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r. \tag{2.34}$$

But

$$\lim_{m=\infty} \operatorname*{Inf}_{\eta^{m}} \operatorname*{Sup}_{\xi} r \geqq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r. \tag{2.35}$$

Hence

$$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r \tag{2.36}$$

Theorem 2.1 is an immediate consequence of (2.36) and Lemma 1.3.

2.5. *Theorems on complete classes of decision functions and minimax solutions.* For any positive ϵ we shall say that the randomized decision function η_0 is an ϵ-Bayes solution relative to the a priori distribution ξ if

$$r(\xi, \eta_0) \leqq \operatorname*{Inf}_{\eta} r(\xi, \eta) + \epsilon. \tag{2.37}$$

If η_0 satisfies (2.37) for $\epsilon = 0$, we shall say that η_0 is a Bayes solution relative to ξ.

A randomized decision rule η_1 is said to be uniformly better than η_2 if

$$r(F, \eta_1) \leqq r(F, \eta_2) \text{ for all } F \tag{2.38}$$

and if

$$r(F, \eta_1) < r(F, \eta_2) \text{ at least for one } F. \tag{2.39}$$

A class C of randomized decision functions η is said to be complete if for any η not in C we can find an element η^* in C such that η^* is uniformly better than η.

THEOREM 2.2. *If Conditions* 2.1–2.5 *are fulfilled, then for any* $\epsilon > 0$ *the class* C_ϵ *of all* ϵ*-Bayes solutions corresponding to all possible a priori distributions* ξ *is a complete class.*

PROOF: Let η_0 be a randomized decision function that is not an ϵ-Bayes solution relative to any ξ. That is,

$$r(\xi, \eta_0) > \operatorname*{Inf}_{\eta} r(\xi, \eta) + \epsilon \text{ for all } \xi. \tag{2.40}$$

If $r(F, \eta_0) = \infty$ for all F, then there is evidently an element of C_ϵ that is uniformly better than η_0. Thus, we can restrict ourselves to the case where

$$r(F, \eta_0) < \infty \text{ at least for one } F. \tag{2.41}$$

Put

$$W^*(F, d) = W(F, d) - r(F, \eta_0) \tag{2.42}$$

and let $r^*(\xi, \eta)$ denote the risk when $W(F, d)$ is replaced by $W^*(F, d)$. Then

$$r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0). \tag{2.43}$$

Let Q^m denote the class of all decision functions $d(x)$ for which $n(x) \leqq m$ identically in x. Furthermore, denote any η for which $\eta(Q^m) = 1$ by η^m. We shall first prove the following relation.

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r^*(\xi, \eta^m) = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*(\xi, \eta^m) \tag{2.44}$$

for any positive integral value m. For any positive constant c, let Ω_c denote the class of all elements F for which $r(F, \eta_0) \leqq c$.

Clearly, Conditions 2.1-2.5 remain valid if we replace $W(F, d)$ by $W^*(F, d)$ and Ω by Ω_c where c is restricted to values for which Ω_c is not empty. Hence, Theorem 2.1 can be applied and we obtain

$$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^*(\xi^c, \eta^m) = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi^c} r^*(\xi^c, \eta^m), \tag{2.45}$$

where ξ^c denotes any ξ for which $\xi(\Omega_c) = 1$. Let h and w be two positive values for which

$$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^*(\xi^c, \eta^m) \geqq -h \quad \text{for all} \quad c \tag{2.46}$$

and

$$r(F, \eta^m) \leqq w \text{ for all } F \text{ and all } \eta^m. \tag{2.47}$$

Clearly, such two constants h and ω exist. From (2.46) and Lemma 1.3 we obtain

$$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*(\xi, \eta^m) \geqq -h. \tag{2.48}$$

Since

$$r^*(F, \eta^m) < -(h + \delta) \text{ for any } F \text{ not in } \Omega_{h+\delta+w} (\delta > 0), \tag{2.49}$$

it follows from (2.48) that

$$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi^c} r^* = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^* \quad \text{for all} \quad c > h + w. \tag{2.50}$$

From (2.45) and (2.50) we obtain

$$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^* = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^* \quad \text{for all} \quad c > h + w. \tag{2.51}$$

Hence,

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r^* \geqq \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*. \tag{2.51a}$$

Because of Lemma 1.3, the equality sign must hold and (2.44) is proved.

Since η_0 is not an element of C_ϵ, we must have

$$\operatorname*{Inf}_{\eta} r(\xi, \eta) < r(\xi, \eta_0) - \epsilon. \tag{2.52}$$

From this it follows that

$$\operatorname*{Inf}_{\eta} r^*(\xi, \eta) \leqq -\epsilon. \tag{2.53}$$

Hence

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^*(\xi, \eta) \leqq -\epsilon. \tag{2.54}$$

It was shown in the proof of Lemma 2.1 that for any $\rho > 0$ there exists a positive integer m_ρ, depending only on ρ, such that

$$\operatorname*{Inf}_{\eta^{m_\rho}} r(\xi, \eta^{m_\rho}) \leqq \operatorname*{Inf}_{\eta} r(\xi, \eta) + \rho \text{ for all } \xi. \tag{2.55}$$

From (2.44), (2.54) and (2.55) it follows that there exists a positive integer m_0, namely $m_0 = m_{\epsilon/2}$, such that

$$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*(\xi, \eta^m) \leq -\frac{\epsilon}{2} \text{ for any } m \geqq m_0. \tag{2.56}$$

From (2.44) and (2.56) it follows that there exists an a priori distribution ξ_1 and an ϵ-Bayes solution η_1^m relative to ξ_1 such that

$$r^*(F, \eta_1^m) \leqq -\frac{\epsilon}{4} \text{ for all } F. \tag{2.57}$$

Hence, because of (2.43),

$$r(F, \eta_1^m) \leqq r(F, \eta_0) - \frac{\epsilon}{4} \text{ for all } F. \tag{2.58}$$

and Theorem 2.2 is proved.

THEOREM 2.3. *If D is compact, and if Conditions* 2.1, 2.2, 2.4, 2.5 *are fulfilled, then there exists a minimax solution, i.e., a decision rule η_0 for which*

$$\operatorname{Sup}_F r(F, \eta_0) \leqq \operatorname{Sup}_F r(F, \eta) \text{ for all } \eta. \tag{2.59}$$

To prove the above theorem, we shall first prove the following lemma.

LEMMA 2.4. *If D is compact and if Conditions 2.1, 2.2, 2.4, 2.5 are fulfilled, then for any sequence $\{\eta_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ of randomized decision functions for which $r(F, \eta_i)$ is a bounded function of F and i, there exists a subsequence $\{\eta_{i_j}\}$ $(j = 1, 2, \cdots, \text{ad inf.})$ and a randomized decision function η_0 such that*

$$\liminf_{j=0} r(\xi, \eta_{i_j}) \geqq r(\xi, \eta_0) \text{ for all } \xi. \tag{2.60}$$

PROOF: Let $\zeta_i = \zeta(\eta_i)$ (defined in Section 2.3) be given by $\{z_{r,i}(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1x^2 \cdots x^r,i}\}$ $(r = 1, 2, \cdots, \text{ad inf.})$. Thus, $z_{r,i}(x^1, \cdots, x^r)$ is the conditional probability that we shall take an observation on X^{r+1} using the rule η_i and knowing that the first r observations are given by $x^1, \cdots, x^r$ and that experimentation was not terminated for $(x^1, \cdots, x^k)$ $(k < r)$. As stated in section 2.3, for any $r, x^1, \cdots, x^r$ the symbol $\delta_{x^1 \cdots x^r,i}$ denotes the conditional probability distribution of the selected d when η_i is used and is known that the first r observations are equal to $x^1, \cdots, x^r$ and that $n(x) = r$. Since there are only countably many finite samples $(x^1, \cdots, x^r)$, it is possible to find a subsequence $\{i_j\}$ of $\{i\}$ such that $\lim_{j=\infty} z_{r,i_j}(x^1, \cdots, x^r)$ and $\lim_{j=\infty} \delta_{x^1 \cdots x^r,i_j}$ exist [5] Put

$$\lim_{j=\infty} z_{r,i_j}(x^1, \cdots, x^r) = z_{r,0}(x^1, \cdots, x^r) \tag{2.61}$$

and

$$\lim_{j=\infty} \delta_{x^1 \cdots x^r,i_j} = \delta_{x^1 \cdots x^r,0}. \tag{2.62}$$

As shown in section 2.3, there exists a randomized decision function η_0 such $\zeta_0 = \zeta(\eta_0)$ is given by $\{z_{r,0}(x^1, \cdots, x^2)\}$ and $\{\delta_{x^1 \cdots x^r,0}\}$. Let $q_{r,i}(x^1, \cdots, x^r \mid \xi)$ denote the probability that the sample $(x^1, \cdots, x^r)$ will be obtained and that experimentation will be stopped at the r-th observation when ξ is the a priori distribution and η_i is the decision rule used by the statistician. For any sample $(x^1, \cdots, x^r)$ let $R_i(x^1, \cdots, x^r)$ denote the expected value of $W(F, d)$ when the distribution of F is equal to the a posteriori distribution of F as implied by ξ and $(x^1, \cdots, x^r)$ and where d is a chance variable independent of F with the probability distribution $\delta_{x^1 \cdots x^r,i}$. Since, $r(\xi, \eta_i)$ is bounded by assumption, the probability that experimentation will go on indefinitely is equal to zero. From this it follows that

$$\sum_{r,x^1 \cdots x^r} q_{r,i}(x^1, \cdots, x^r \mid \xi) = 1 \text{ for all } \xi. \tag{2.63}$$

[5] The existence of $\lim_{j=\infty} \delta_{x^1 \cdots x^r,i_j}$ follows from the compactness of D (see Theorem 3.6 in [3]).

Then $r(\xi, \eta_i)$ is given by

$$r(\xi, \eta_i)$$

$$= \sum_{r, x^1, \cdots, x^r} q_{r,i}(x^1, \cdots, x^r \mid \xi) \left[R_i(x^1, \cdots, x^r) + \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r, \mathfrak{D})\, d\eta_i}{\int_{Q_{x^1 \cdots x^r}} d\eta_i} \right] \tag{2.64}$$

where $Q_{x^1 \cdots x^r}$ is the totality of all decision functions $d(x)$ for which $n(y) = r$ whenever $y^1 = x^1, \cdots, y^r = x^r$. Clearly,

$$\lim_{j=\infty} q_{r,i_j}(x^1, \cdots, x^r \mid \xi) = q_{r,0}(x^1, \cdots, x^r \mid \xi). \tag{2.65}$$

Since D is compact and since $W(F, d)$ is a continuous function of d uniformly in F (in the sense of the metric defined in D), we have

$$\lim_{j=\infty} R_{i_j}(x^1, \cdots, x^r) = R_0(x^1, \cdots, x^r). \tag{2.66}$$

From Condition 2.5 it follows that

$$\lim_{j=\infty} \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r, \mathfrak{D})\, d\eta_{i_j}}{\int_{Q_{x^1 \cdots x^r}} d\eta_{i_j}} = \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r \mathfrak{D}_{,0})\, d\eta_0}{\int_{Q_{x^1 \cdots x^r}} d\eta_0}. \tag{2.67}$$

Lemma 2.4 is an immediate consequence of the equations (2.64) — (2.67).

We are now in a position to prove Theorem 2.3. Because of Theorem 2.1 there exists a sequence $\{\eta_i\}$ such that

$$\lim_{i=\infty} \operatorname*{Sup}_F r(F, \eta_i) = \operatorname*{Inf}_\eta \operatorname*{Sup}_F r(F, \eta). \tag{2.68}$$

According to Lemma 2.4 there exists a subsequence $\{\eta_{i_j}\}$ ($j = 1, 2, \cdots$, ad inf.) and a randomized decision function η_0 such that

$$\liminf_{j=\infty} r(F, \eta_{i_j}) \geqq r(F, \eta_0) \text{ for all } F. \tag{2.69}$$

It follows from (2.68) and (2.69) that η_0 is a minimax solution and Theorem 2.3 is proved.

THEOREM 2.4. *If D is compact and if Conditions* 2.1, 2.2, 2.4, 2.5 *are fulfilled, then for any ξ there exists a Bayes solution relative to ξ.*

This theorem is an immediate consequence of Lemma 2.4.

We shall say that η_0 is a Bayes solution in the wide sense, if there exists a sequence $\{\xi_i\}$ ($i = 1, 2, \cdots$, ad inf.) such that

$$\lim_{i=\infty} [r(\xi_i, \eta_0) - \operatorname*{Inf}_\eta r(\xi_i, \eta)] = 0. \tag{2.70}$$

We shall say that η_0 is a Bayes solution in the strict sense, if there exists a ξ such that η_0 is a Bayes solution relative to ξ.

THEOREM 2.5. *If D is compact and Conditions* 2.1–2.5 *hold, then the class of all Bayes solutions in the wide sense is a complete class.*

PROOF: Let η_0 be a decision rule that is not a Bayes solution in the wide sense. Consider the weight function $W^*(F, d) = W(F, d) - r(F, \eta_0)$. We may assume that $r(F, \eta_0) < \infty$ for at least some F, since otherwise there obviously exists a Bayes solution in the wide sense that is uniformly better than η_0. Then it follows easily from (2.44) and Lemmas 2.1 and 1.3 that

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^*(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r^*(\xi, \eta) = v^* \quad \text{(say)}, \tag{2.71}$$

where $r^*(\xi, \eta)$ is the risk corresponding to $W^*(F, d)$, i.e.,

$$r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0). \tag{2.72}$$

Theorem 2.3 is clearly applicable to the risk function $r^*(\xi, \eta)$. Then, there exists a minimax solution η_1 for the problem corresponding to the new weight function $W^*(F, d)$. Since, because of 2.72, $v^* \leqq 0$, we have

$$r^*(\xi, \eta_1) = r(\xi, \eta_1) - r(\xi, \eta_0) \leqq 0 \text{ for all } \xi. \tag{2.73}$$

Our theorem is proved, if we can show that η_1 is a Bayes solution in the wide sense. Let $\{\xi_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of a priori distributions such that

$$\lim_{i=\infty} \operatorname{Inf}_{\eta} r^*(\xi_i, \eta) = v^*. \tag{2.74}$$

Since η_1 is a minimax solution, we must have

$$r^*(\xi_i, \eta_1) \leqq v^*. \tag{2.75}$$

It follows from (2.74) and (2.75) that η_1 is a Bayes solution in the wide sense and our theorem is proved.

We shall now formulate an additional condition which will permit the derivation of some stronger theorems. First, we shall give a convergence definition in the space Ω. We shall say that F_i converges to F in the ordinary sense if

$$\lim_{i=\infty} p_r(x^1, \cdots, x^r | F_i) = p_r(x^1, \cdots, x^r | F) \qquad (r = 1, 2, \cdots, \text{ad inf.}). \tag{2.76}$$

Here $p_r(x^1, \cdots, x^r \mid F)$ denotes the probability, under F, that the first r observations will be equal to $x^1, \cdots, x^r$, respectively. We shall say that a subset ω of Ω is compact in the ordinary sense, if ω is compact in the sense of the convergence definition (2.76).

CONDITION 2.6. *The space Ω is compact in the ordinary sense. If F_i converges to F, as $i \to \infty$, in the ordinary sense, then*

$$\lim_{i=\infty} W(F_i, d) = W(F, d)$$

uniformly in d.

THEOREM 2.6. *If D is compact and if Conditions* 2 1, 2.2, 2.4, 2.5, 2.6 *hold, then:*

(i) *there exists a least favorable a priori distribution, i.e., an a priori distribution* ξ_0 *for which*

$$\operatorname*{Inf}_{\eta} r(\xi_0, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta).$$

(ii) *A minimax solution exists and any minimax solution is a Bayes solution in the strict sense.*

(iii) *If* η_0 *is a decision rule which is not a Bayes solution in the strict sense and for which* $r(F, \eta_0)$ *is a bounded function of* F, *then there exists a decision rule* η_1 *which is a Bayes solution in the strict sense and is uniformly better than* η_0.

PROOF: Let $\{\xi_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of a priori distributions such that

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta} r(\xi_i, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta). \tag{2.77}$$

Since Ω is compact in the ordinary sense, there exists an a priori distribution ξ_0 and a subsequence $\{\xi_{i_j}\}$ or $\{\xi_i\}$ such that

$$\lim_{j=\infty} \xi_{i_j}(\omega) = \xi_0(\omega) \tag{2.78}$$

for any subset ω of Ω which is open (in the sense of the ordinary convergence definition in Ω) and for which $\xi_0(\omega^*) = 0$, where ω^* denotes the set of all boundary points of ω. We shall show that ξ_0 is a least favorable distribution. Assume that it is not. Then there exists a decision function $\mathfrak{D}_0 = d_0(x)$ such that

$$r(\xi_0, \mathfrak{D}_0) \leqq v - \delta, \tag{2.79}$$

where $\delta > 0$ and v denotes the common value of $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r$ and $\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r$. It was shown in the proof of Lemma 2.1 that (2.79) implies the existence of a decision function $\mathfrak{D}_1 = d_1(x)$ and that of a positive integer m such that

$$n(x; \mathfrak{D}_1) \leqq m \text{ for all } x \tag{2.80}$$

and

$$r(\xi_0, \mathfrak{D}_1) \leqq v - \frac{\delta}{2}. \tag{2.81}$$

Since $c(x^1, \cdots, x^n, \mathfrak{D}_1)$ and $W(F, d)$ are uniformly bounded and $W(F, d)$ is continuous in F uniformly in d, we have

$$\lim_{i=\infty} r \quad (F_i, \mathfrak{D}_1) = r(F, \mathfrak{D}_1) \tag{2.82}$$

for any sequence $\{F_i\}$ for which $F_i \to F$ in the ordinary sense. From (2.78), (2.82) and the compactness of Ω (in the ordinary sense) it follows that

$$\lim_{j=\infty} r \quad (\xi_{i_j}, \mathfrak{D}_1) = r(\xi_0, \mathfrak{D}_1) \leqq v - \frac{\delta}{2}. \tag{2.83}$$

But this is in contradiction to (2.77) and, therefore, ξ_0 must be a least favorable distribution. Hence, statement (i) of our theorem is proved.

Statement (ii) is an immediate consequence of Theorems (2.1), (2.3) and statement (i) of Theorem (2.6).

To prove (iii), replace the weight function $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \eta_0)$ where η_0 satisfies the conditions imposed on it in (iii).

We shall show that (i) remains valid also when $W(F, d)$ is replaced by $W^*(F, d)$. This is not clear, since $W^*(F, d)$ may not be continuous in F. First we shall prove that

$$\liminf_{i=\infty} r(\xi_i', \eta_0) \geqq r(\xi_0', \eta_0) \tag{2.84}$$

for any sequence $\{\xi_i'\}$ for which $\xi_i' \to \xi_0'$ in the ordinary sense, i.e., for which

$$\lim_{i=\infty} \xi_i'(\omega) = \xi_0'(\omega) \tag{2.85}$$

for any open subset ω (open in the sense of ordinary convergence defined in Ω) whose boundary has probability measure zero according to ξ_0'. For any sample $x^1, \cdots, x^r$ let $q_{ri}(x^1, \cdots, x^r)$ denote the probability that the first r observations will be equal to $x^1, \cdots, x^r$, respectively, when ξ_i' is the a priori distribution. Clearly,

$$q_{ri}(x^1, \cdots, x^r) = \int_\Omega p_r(x^1, \cdots, x^r \mid F)\, d\xi_i'. \tag{2.86}$$

Since $p_r(x^1, \cdots, x^r \mid F)$ is a continuous function of F, we have

$$\lim_{i=\infty} q_{ri}(x^1, \cdots, x^r) = q_{r0}(x^1, \cdots, x^r). \tag{2.87}$$

The function $r(\xi, \eta_0)$ can be split into two parts, i.e., $r(\xi, \eta_0) = r_1(\xi, \eta_0) + r_2(\xi, \eta_0)$ where r_1 is the expected value of the loss $W(F, d)$ and r_2 is the expected cost of experimentation. Since $W(F, d)$ is a bounded function of F and d, and since $W(F, d)$ is continuous in F uniformly in d, we have

$$\lim r_1(\xi_i', \eta_0) = r_1(\xi_0', \eta_0) \tag{2.88}$$

for any sequence $\{\xi_i'\}$ which satisfies (2.85). To prove (2.84), we merely have to show that

$$\liminf_{i=\infty} r_2(\xi_i', \eta_0) \geqq r_2(\xi_0', \eta_0). \tag{2.89}$$

But

$$r_2(\xi_i', \eta_0) = \sum_{r, x^1, \cdots, x^r} q_{ri}(x^1, \cdots, x^r) \int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r; \mathfrak{D})\, d\eta_0 \tag{2.90}$$

where $Q_{x^1 \cdots x^r}$ is the totality of all decision functions $d(x)$ with the property that $d(y) = r$ for any y whose first r coordinates are equal to $x^1, \cdots, x^r$, respectively. Equation (2.89) is an immediate consequence of (2.87) and (2.90). Hence, (2.84) is proved.

Let $r^*(\xi, \eta)$ be the risk function when $W(F, d)$ is replaced by $W^*(F, d)$, i.e., $r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0)$. Let, furthermore, $\{\xi_i^*\}$ be a sequence of a priori distributions such that

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta} r^*(\xi_i^*, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^*(\xi, \eta). \tag{2.91}$$

There exists a subsequent $\{\xi_{i_j}^*\}$ of the sequence $\{\xi_i^*\}$ such that $\xi_{i_j}^*$ converges (in the ordinary sense) to a limit distribution ξ_0^* as $j \to \infty$. We shall show that ξ_0^* is a least favorable distribution. For suppose that ξ_0^* is not a least favorable distribution. Then there exists a decision function $\mathfrak{D}_0^* = d_0^*(x)$ such that

$$r^*(\xi_0^*, \mathfrak{D}_0^*) \leqq v^* - \delta \tag{2.92}$$

where $\delta > 0$ and $v^* = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^* = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r^*$. But then there exists a decision function $\mathfrak{D}_1^* = d_1^*(x)$ and a positive integer m such that

$$n(x; \mathfrak{D}_1^*) \leqq m \text{ for all } x \tag{2.93}$$

and

$$r^*(\xi_0^*, \mathfrak{D}_1^*) \leqq v^* - \frac{\delta}{2}. \tag{2.94}$$

Since $r^*(\xi, \mathfrak{D}_1^*) = r(\xi, \mathfrak{D}_1^*) - r(\xi, \eta_0)$, and since

$$\lim_{j=\infty} r(\xi_{i_j}^*, \mathfrak{D}_1^*) = r(\xi_0^*, \mathfrak{D}_1^*),$$

it follows from (2.84) and (2.94) that

$$\limsup_{j=\infty} r^*(\xi_{i_j}^*, \mathfrak{D}_1^*) \leqq v^* - \frac{\delta}{2} \tag{2.95}$$

which is in contradiction to (2.91). Hence, the validity of (i) is proved also when $W(F, d)$ is replaced by $W^*(F, d)$. Clearly, also (ii) remains valid when $W(F, d)$ is replaced by $W^*(F, d)$.

Let η_1 be a minimax solution relative to the problem corresponding to $W^*(F, d)$. Then because of (ii), η_1 is a Bayes solution in the strict sense. Since η_0 is not a Bayes solution in the strict sense, $\eta_1 \neq \eta_0$ and $v^* < 0$. Hence η_1 is uniformly better than η_0. This completes the proof of Theorem 2.6.

We shall now replace Condition 2.6 by the following weaker one.

CONDITION 2.6*. *There exists a sequence* $\{\Omega_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *of subsets of* Ω *such that Condition* 2.6 *is fulfilled when* Ω *is replaced by* Ω_i, $\Omega_{i+1} \supset \Omega_i$ *and* $\lim_{i=\infty} \Omega_i = \Omega$.

We shall say that η_i converges weakly to η as $i \to \infty$, if $\lim_{i=\infty} \zeta(\eta_i) = \zeta(\eta)$. We shall also say that η is a weak limit of η_i. This limit definition seems to be natural, since $r(\xi, \eta_1) = r(\xi, \eta_2)$ if $\zeta(\eta_2) = \zeta(\eta_1)$. We shall now prove the following theorem:

THEOREM 2.7. *If D is compact and if Conditions* 2.1, 2 2, 2.4, 2.5 *and* 2.6* *are fulfilled, then:*

(i) *A minimax solution exists that is a weak limit of a sequence of Bayes solutions in the strict sense.*

(ii) *Let η_0 be a decision rule for which $r(F, \eta_0)$ is a bounded function of F. Then there exists a decision rule η_1 that is a weak limit of a sequence of Bayes solutions in the strict sense and such that $r(F, \eta_1) \leqq r(F, \eta_0)$ for all F in Ω.*

PROOF: According to theorem 2 6, there exists a decision rule η_i that is a Bayes solution in the strict sense and a minimax solution if Ω is replaced by Ω_i. There exists a subsequence $\{\eta_{i_j}\}$ ($j = 1, 2, \cdots$, ad inf.) of the sequence $\{\eta_i\}$ such that $\{\eta_{i_j}\}$ admits a weak limit. Let η_0 be a weak limit of $\{\eta_{i_j}\}$. Then, as shown in the proof of Lemma 2.4, equation (2 60) holds and η_0 is a minimax solution relative to the original space Ω. Thus, statement (i) is proved.

To prove (ii), replace $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \eta_0)$. According to Theorem 2.6 there exists a decision rule η_{1i} such that η_{1i} is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i and $W(F, d)$ by $W^*(F, d)$ Clearly, η_{1i} remains to be a Bayes solution in the strict sense also relative to Ω and $W(F, d)$. Since η_{1i} is a minimax solution relative to Ω_i and $W^*(F, d)$, we have

$$r(F, \eta_{1i}) \leqq r(F, \eta_0) \text{ for all } F \text{ in } \Omega_i. \tag{2.96}$$

Let $\{\eta_{1i_j}\}$ be a subsequence of the sequence $\{\eta_{1i}\}$ such that $\{\eta_{1i_j}\}$ admits a weak limit η_1. Then, (2.60) holds for $\{\eta_{1i_j}\}$ and η_1, and

$$r(F, \eta_1) \leqq r(F, \eta_0) \text{ for all } F \text{ in } \Omega. \tag{2.97}$$

Since η_1 is a weak limit of strict Bayes solution, statement (ii) is proved.

3. Statistical decision functions: the case of continuous chance variables.

3.1. *Introductory remarks.* In this section we shall be concerned with the case where the probability distribution F of X is absolutely continuous, i.e., for any element F of Ω and for any positive integer r there exists a joint density function $p_r(x^1, \cdots, x^r \mid F)$ of the first r chance variables $X^1, \cdots, X^r$.

The continuous case can immediately be reduced to the discrete case discussed in section 2 if the observations are not given exactly but only up to a finite number of decimal places. More precisely, we mean this: For each i, let the real axis R be subdivided into a denumerable number of disjoint sets $R_{i1}, R_{i2}, \cdots$, ad inf. Suppose that the observed value x^i of X^i is not given exactly; it is merely known which element of the sequence $\{R_{ij}\}$ ($j = 1, 2, \cdots$, ad inf.) contains x^i. This is the situation, for example, if the value of x^i is given merely up to a finite number, say r, decimal places (r fixed, independent of i). This case can be reduced to the previously discussed discrete case, since we can regard the sets R_{ij} as our points, i.e, we can replace the chance variable X^i by Y^i where Y^i can take only the values $R_{i1}, R_{i2}, \cdots$, ad inf. (Y^i takes the value R_{ij} if X^i falls in R_{ij}). If $W(F_1, d) = W(F_2, d)$ whenever the distribution of Y under

F_1 is identical with that under F_2, only the chance variables Y^1, Y^2, $\cdots$, etc play a role in the decision problem and we have the discrete case. If, the latter condition on the weight function is not fulfilled, i e., if there exists a pair (F_1, F_2) such that $W(F_1, d) \neq W(F_2, d)$ for some d and the distribution of Y is the same under F_1 as under F_2, we can still reduce the problem to the discrete case, if in the discrete case we permit the weight W to depend also on a third extraneous variable G, i e., if we put $W = W(F, G, d)$, where G is a variable about whose value the sample does not give any information. The results obtained in the discrete case can easily be generalized to include the situation where $W = W(F, G, d)$.

In practical applications the observed value x^i of X^i will usually be given only up to a certain number of decimal places and, thus, the problem can be reduced to the discrete case. Nevertheless, it seems desirable from the theoretical point of view to develop the theory of the continuous case, assuming that the observed value x^i of X^i is given precisely.

In section 2.3 an alternative definition of a randomized decision rule was given in terms of two sequences of functions $\{z_r(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r}\}$ ($r = 1, 2, \cdots$, ad inf.). We used the symbol ζ to denote a randomized decision rule given by two such sequences. It was shown in the discrete case that the use of a randomized decision function η generates a certain $\zeta = \zeta(\eta)$, and that for any given ζ there exists an η such that $\zeta = \zeta(\eta)$. Furthermore, because of Condition 2 5, in the discrete case we had $r(F, \eta_1) = r(F, \eta_2)$ if $\zeta(\eta_1) = \zeta(\eta_2)$. It would be possible to develop a similar theory as to the relation between ζ and η also in the continuous case However, a somewhat different procedure will be followed for the sake of simplicity. Instead of the decision functions $d(x)$, we shall regard the ζ's as the pure strategies of the statistician, i.e., we replace the space Q of all decision functions $d(x)$ by the space Z of all randomized decisions rules ζ. It will then be necessary to consider probability measures η defined over an additive class of subsets of Z. It will be sufficient, as will be seen later, to consider only discrete probability measures η. A probability measure η is said to be discrete, if it assigns the probability 1 to some denumerable subset of Z. Any discrete η will clearly generate a certain $\zeta = \zeta(\eta)$. In the next section we shall formulate some conditions which will imply that $r(F, \eta_1) = r(F, \eta_2)$ if $\zeta(\eta_1) = \zeta(\eta_2)$. Thus, it will be possible to restrict ourselves to consideration of pure strategies ζ which will cause considerable simplifications.

The definitions of various notions given in the discrete case, such as minimax solution, Bayes solution, a priori distribution ξ in Ω, least favorable a priori distribution, complete class of decision functions, etc. can immediately be extended to the continuous case and will, therefore, not be restated here.

3.2 *Conditions on Ω, D, $W(F, d)$ and the cost function.* In this section we shall formulate conditions similar to those given in the discrete case.

CONDITION 3.1. *Each element F of Ω is absolutely continuous.*

CONDITION 3 2. *$W(F, d)$ is a bounded function of F and d.*

CONDITION 3.3. *The space D is compact in the sense of its intrinsic metric $\delta(d_1, d_2; \Omega)$ (see equation 2.2).*

This condition is somewhat stronger than the corresponding Condition 2.3. While it may be possible to weaken this condition, it would make the proofs of certain theorems considerably more involved.

CONDITION 3.4. *The cost of experimentation* $c(x^1, \cdots, x^m)$ *does not depend on* ζ. *It is non-negative and* $\lim_{m=\infty} c(x^1, \cdots, x^m) = \infty$ *uniformly in* $x^1, \cdots, x^m$. *For each positive integral value* m, $c(x^1, \cdots, x^m)$ *is a bounded function of* $x^1, \cdots, x^m$.

This condition is stronger than Conditions 2.4 and 2.5 postulated in the discrete case The reason for formulating a stronger condition here is that we wish the relation $r(F, \eta_1) = r(F, \eta_2)$ to be fulfilled whenever $\zeta(\eta_1) = \zeta(\eta_2)$ which will make it possible for us to eliminate the consideration of η's altogether. Since the ζ's are regarded here as the pure strategies of the statistician, it is not clear what kind of dependence of the cost on ζ would be consistent with the requirement that $r(F, \eta_1) = r(F, \eta_2)$ whenever $\zeta(\eta_1) = \zeta(\eta_2)$.

We shall say that $F_i \longrightarrow F$ in the ordinary sense, if for any positive integral value m

$$\lim_{i=\infty} \int_{S_m} p_m(x^1, \cdots, x^m \mid F_i)\, dx^1 \cdots dx^m = \int_{S_m} p_m(x^1, \cdots, x^m \mid F)\, dx^1 \cdots dx^m$$

uniformly in S_m where S_m is a subset of the m-dimensional sample space.

CONDITION 3.5. *The space* Ω *is separable in the sense of the above convergence definition.*[6]

No such condition was formulated in the discrete case for the simple reason that in the discrete case Ω is always separable in the sense of the convergence definition given in (2.76).

3.3. *Some lemmas.* We shall first give a convergence definition in the space Z of all ζ's which is somewhat different from the one given in the discrete case. Let $h_r(x^1, x^2, \cdots, x^r, D^*)$ denote the probability that experimentation will be terminated with the rth observation and that the final decision d selected will be an element of D^*, knowing that the first r observations are equal to $x^1, \cdots, x^r$, respectively. That is,

$$\begin{aligned} h_r(x^1, \cdots, x^r, D^*) &= z_1(x^1) z_2(x^1, x^2) \\ &\quad \cdots z_{r-1}(x^1, \cdots, x^{r-1})(1 - z_r(x^1, \cdots, x^r))\delta_{x^1 \cdots x^r}(D^*). \end{aligned} \tag{3.1}$$

Clearly, the functions $h_r(x^1, \cdots, x^r, D^*)$ are non-negative and satisfy the following conditions:

$$\sum_{r=1}^{m} h_r(x^1, \cdots, x^r, D^*) \leqq 1 \text{ for any } D^* \text{ and for any sample } (x^1, \cdots, x^m). \tag{3.2}$$

$$\sum_{j=1}^{\infty} h_r(x^1, \cdots, x^r, D_j^*) = h_r(x^1, \cdots, x^r, D^*), \tag{3.3}$$

if $\sum_{j=1}^{\infty} D_j^* = D^*$ and $D_1^*, D_2^*, \cdots$, etc. are disjoint.

[6] For a definition of a separable space, see F. Hausdorff, *Mengenlehre* (3rd edition), p. 125.

One can easily verify that for any sequence of non-negative functions $\{h_r(x^1, \cdots, x^r, D^*)\}$ $(r = 1, 2, \cdots)$ satisfying (3.2) and (3.3) there exists exactly one sequence $\{z_r(x^1, \cdots, x^r)\}$ and one sequence $\{\delta_{x^1 \cdots x^r}(D^*)\}$ such that (3.1) is fulfilled. Thus, a randomized decision rule ζ can be given by a sequence $\{h_r(x^1, \cdots, x^r, D^*)\}$ satisfying (3.2) and (3.3). The functions $z_r(x^1, \cdots, x^r)$ and $\delta_{x^1 \cdots x^r}$ need be defined only for samples $x^1, \cdots, x^r$ for which $z_i(x^1, \cdots, x^i) > 0$ for $i = 1, \cdots, r-1$. The above mentioned uniqueness of $z_r(x^1, \cdots, x^r)$ and $\delta_{x^1 \cdots x^r}$ was meant to hold if the definition of these functions is restricted to such samples $x^1, \cdots, x^r$.

For any bounded subset S_r of the r-dimensional sample space, let

$$H_r(S_r, D^*) = \int_{S_r} h_r(x^1, \cdots, x^r, D^*)\, dx^1 \cdots dx^r. \tag{3.4}$$

Let $\{\zeta_i\}(i = 0, 1, 2, \cdots,$ ad inf.) be a sequence of decision rules, and $H_{r,i}(S_r, D^*)$ be the function $H_r(S_r, D^*)$ corresponding to ζ_i. We shall say that

$$\lim_{i=\infty} \zeta_i = \zeta_0 \tag{3.5}$$

if

$$\lim_{i=\infty} H_{r,i}(S_r, D^*) = H_{r,0}(S_r, D^*) \tag{3.6}$$

for any r, any bounded set S_r and for any D^* that is an element of a sequence $\{D_{k_1 \cdots k_l}\}$ $(k_j = 1, \cdots, r_j;\ j = 1, \cdots, l;\ l = 1, 2, \cdots,$ ad inf.) of subsets of D satisfying the following conditions:

$$\sum_{k_1=1}^{r_1} D_{k_1} = D;\ \sum_{k_l} D_{k_1 \cdots k_l} = D_{k_1 k_2 \cdots k_{l-1}}, \tag{3.7}$$

$$D_{k_1 \cdots k_{l-1} 1}, \cdots, D_{k_1 \cdots k_{l-1} r_l} \text{ are disjoint,} \tag{3.8}$$

and

(3.9) the diameter of $D_{k_1 \cdots k_l}$ converges to zero as $l \to \infty$ uniformly in $k_1, \cdots, k_l$.

LEMMA 3.1. *For any sequence $\{\zeta_i\}(i = 1, 2, \cdots,$ ad inf.) of decision rules there exists a subsequence $\{\zeta_{i_j}\}$ $(j = 1, 2, \cdots,$ ad inf.) and a decision rule ζ_0 such that $\lim_{j=\infty} \zeta_{i_j} = \zeta_0$.*

PROOF: Let $H_{r,i}(S_r, D^*)$ $(r = 1, 2, \cdots,$ ad inf.) be the sequence of functions associated with ζ_i. Let, furthermore, $\{D^*_{k_1 \cdots k_l}\}$ be a sequence of subsets of D satisfying the relations (3.7), (3.8) and (3.9). Clearly, for any fixed r and any fixed element $D^*_{k_1 \cdots k_l}$ of the sequence $\{D^*_{k_1 \cdots k_l}\}$, it is possible to find a subsequence $\{i_j\}$ $(j = 1, 2, \cdots,$ ad inf.) of the sequence $\{i\}$ (the subsequence $\{i_j\}$ may depend on r and $D^*_{k_1 \cdots k_l}$) and a set function $H_{r,0}(S_r)$ such that

$$\lim_{j=\infty} H_{r,i_j}(S_r, D^*_{k_1 \cdots k_l}) = H_{r,0}(S_r). \tag{3.10}$$

Using the well known diagonal procedure, it is therefore possible to find a fixed

subsequence $\{i_j\}$ (independent of r and D^*) and a sequence of set functions $\{H_{r,0}(S_r, D^*_{k_1 \cdots k_l})\}$ such that

$$\lim_{j=\infty} H_{r,i_j}(S_r, D^*_{k_1 \cdots k_l}) = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}) \tag{3.11}$$

for all values of $r, k_1, \cdots, k_l$ and l.

To complete the proof of Lemma 3.1, it remains to be shown that there exists a decision rule ζ_0 such that the associated function $H_r(S_r, D^*)$ is equal to $H_{r,0}(S_r, D^*)$ for any D^* that is an element of $\{D^*_{k_1 \cdots k_l}\}$. Since $h_{r,i}(x^1, \cdots, x^r, D^*)$ is uniformly bounded, the set function $H_{r,0}(S_r, D^*_{k_1 \cdots k_l})$ is absolutely continuous. Hence for any values of $k_1, \cdots, k_l$ there exists a function $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ such that

$$\int_{S_r} h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l}) dx^1 \cdots dx^r = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}). \tag{3.12}$$

The existence of a ζ_0 with the desired property is proved, if we show that the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ satisfy the relations (3.2) and (3.3). Let $h_r(x^1, \cdots, x^m, D^*) = h_r(x^1, \cdots, x^r, D^*)$ for any $m > r$. Then, since the functions $h_{r,i}$ satisfy (3.2), we have

$$\sum_{r=1}^{m} H_{r,i}(S_m, D^*) \leqq V(S_m) \tag{3.13}$$

where $V(S_m)$ denotes the m-dimensional Lebesgue measure of S_m. From (3.13) it follows that

$$\sum_{r=1}^{m} H_{r,0}(S_m, D^*_{k_1 \cdots k_l}) \leqq V(S_m). \tag{3.14}$$

Hence, the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ must satisfy (3.2) except perhaps on a set of Lebesgue measure zero. Since the functions $h_{r,i}(x^1, \cdots, x^r, D^*)$ satisfy (3.3), we must have

$$H_{r,i}(S_r, D^*_{k_1 \cdots k_{l-1}}) = \sum_{k_l=1}^{r1} H_{r,0}(S_r, D^*_{k_1 \cdots k_l}). \tag{3.15}$$

Hence, the same relation must hold also for $H_{r,0}(S_r, D^*_{k_1 \cdots k_l})$. But this implies that the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ satisfy (3.3) except perhaps on a set of Lebesgue measure zero, and the proof of Lemma 3.1 is completed.

LEMMA 3.2. *Let $T_i(S)$ $(i = 0, 1, 2, \cdots)$ be a non-negative, completely additive set function defined for all measurable subsets S of the r-dimensional sample space M_r. Assume that*

$$T_i(S) \leqq V(S) \tag{3.16}$$

for all S $(i = 0, 1, 2, \cdots,$ ad inf.$)$ where $V(S)$ denotes the Lebesgue measure of S. Let, furthermore, $g(x^1, \cdots, x^r)$ be a non-negative function such that

$$\int_{M_r} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r < \infty, \tag{3.17}$$

If

$$\lim_{i=\infty} T_i(S) = T_0(S) \tag{3.18}$$

then

$$\lim_{i=\infty} \int_{M_r} g(x^1, \cdots, x^r)\, dT_i = \int_{M_r} g(x^1, \cdots, x^r)\, dT_0 . \tag{3.19}$$

PROOF: Let $M_{r,c}$ be the sphere in M_r with center at the origin and radius c. Clearly,

$$\lim_{c=\infty} \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r = \int_{M_r} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r . \tag{3.21}$$

Hence, because of (3.16), we have

$$\lim_{c=\infty} \left[\int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_i - \int_{M_r} g(x^1, \cdots, x^r)\, dT_i \right] = 0 \tag{3.21}$$

uniformly in i. Hence our lemma is proved if we show that

$$\lim_{i=\infty} \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_i = \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_0 \tag{3.22}$$

for any finite c. Let $g_A(x^1, \cdots, x^r) = g(x^1, \cdots, x^r)$ when $g(x^1, \cdots, x^r) \leqq A$, and $= 0$ otherwise. Since

$$\lim_{A=\infty} \int_{M_{r,c}} (g - g_A)\, dx^1 \cdots dx^r = 0$$

it follows from (3.16) that

$$\lim_{A=\infty} \int_{M_{r,c}} (g - g_A)\, dT_i = 0 \tag{3.23}$$

uniformly in i. Hence, our lemma is proved if we can show that

$$\lim_{i=\infty} \int_{M_{r,c}} g_A\, dT_i = \int_{M_{r,c}} g_A\, dT_0 \tag{3.24}$$

for any $c > 0$ and any $A > 0$. Let S_j denote the set of all points in $M_{r,c}$ for which

$$(j - 1)\,\epsilon \leqq g_A < j\,\epsilon \tag{3.25}$$

where ϵ is a given positive number. We have

$$\sum_j (j - 1)\,\epsilon \int_{S_j} dT_i \leqq \int_{M_{r,c}} g_A\, dT_i \leqq \sum_j j\,\epsilon \int_{S_j} dT_i, \quad (i = 0, 1, 2, \cdots). \tag{3.26}$$

Since for any ϵ, j can take only a finite number of values, and since ϵ can be chosen arbitrarily small, our lemma follows easily from (3.18) and (3.26).

LEMMA 3 3. *Let* $\{\zeta_i\}$ *be a sequence of decision rules such that* $\lim_{i=\infty} \zeta_i = \zeta_0$ *and*

$r(F, \zeta_i)$ is a bounded function of F and i ($i \geqq 1$). Then

$$\liminf_{i=\infty} r(F, \zeta_i) \geqq r(F, \zeta_0). \tag{3.27}$$

PROOF: First we shall show that it is sufficient to prove Lemma 3.3 for any finite space D. For this purpose, assume that Lemma 3 3 is true for any finite decision space, but there exists a non-finite compact decision space D and a sequence $\{\zeta_i\}$ such that $\lim_{i=\infty} \zeta_i = \zeta_0$ and

$$\liminf_{i=\infty} r(F, \zeta_i) = r(F, \zeta_0) - \delta \text{ for some } F(\delta > 0). \tag{3.28}$$

Since $\zeta_i \to \zeta_0$, there exists a sequence $\{D_{k_1 \cdots k_l}\}$ of subsets of D satisfying the conditions (3.7)–(3.9) and such that

$$\lim_{i=\infty} H_{r,i}(S_r, D^*_{k_1 \cdots k_l}) = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}) \tag{3.29}$$

where $H_{r,i}(S_r, D^*)$ is the function H_r associated with $\zeta_i (i = 0, 1, 2, \cdots)$. Let λ be a fixed value of l and consider the corresponding finite sequence $\{D_{k_1 \cdots k_\lambda}\}$ of subsets of D. Let k be the number of elements in this finite sequence. We select one point from each element of the finite sequence $\{D_{k_1 \cdots k_\lambda}\}$. Let the points selected be $d_1, d_2, \cdots, d_k$ and let $\bar{D}$ denote the set consisting of the points $d_1, \cdots, d_k$. Let $\bar{\zeta}_i$ be the decision rule defined as follows: the function $h_r(x^1, \cdots, x^r, d_j)$ associated with $\bar{\zeta}_i$ is equal to $h_{r,i}(x^1, \cdots, x^r, D^*_j)$ where D^*_j is equal to the element of the finite sequence $\{D_{k_1 \cdots k_\lambda}\}$ which contains the point $d_j (j = 1, \cdots, k)$. Clearly, because of (3.29),

$$\lim_{i=\infty} \bar{\zeta}_i = \bar{\zeta}_0. \tag{3.30}$$

Furthermore, for sufficiently large λ we obviously have

$$| r(F, \zeta_i) - r(F, \bar{\zeta}_i) | \leqq \epsilon \text{ for } i = 0, 1, 2, \cdots, \text{ad inf.} \tag{3.31}$$

Since for finite D our lemma is assumed to be true, we have

$$\liminf_{i=\infty} r(F, \bar{\zeta}_i) \geqq r(F, \bar{\zeta}_0). \tag{3.32}$$

Choosing $\epsilon \leqq \frac{\delta}{3}$, we obtain a contradiction from (3.28), (3.31) and (3.32). Thus, it is sufficient to prove Lemma 3.3 for finite D. In the remainder of the proof we shall assume that D consists of the points $d_1, \cdots, d_k$.

The probability that we shall take exactly m observations when ζ_i is used and F is true is given by

$$\begin{aligned} &\text{prob. } \{n = m \mid F, \zeta_i\} \\ &= \int_{M_m} p_m(x^1, \cdots, x^m \mid F) h_{m,i}(x^1, \cdots, x^m, D) dx^1, \cdots dx^m \end{aligned} \tag{3.33}$$

where M_m denotes the m-dimensional sample space. Since

$$\lim_{i=\infty} H_{m,i}(S_m, D) = H_{m,0}(S_m, D),$$

it follows from Lemma 3.2 that

$$\lim_{i=\infty} \text{prob}\{n = m \mid F, \zeta_i\} = \text{prob}\{n = m \mid F, \zeta_0\}. \tag{3.34}$$

Hence

$$\lim_{i=\infty} \text{prob}\{n \leqq m \mid F, \zeta_i\} = \text{prob}\{n \leqq m \mid F, \zeta_0\}. \tag{3.35}$$

Since $r(F, \zeta_i)$ is a bounded function of F and i ($i \geqq 1$), we must have

$$\lim_{m=\infty} \text{prob}\{n \leqq m \mid F, \zeta_i\} = 1 \qquad (i = 1, 2, \cdots) \tag{3.36}$$

uniformly in F and i. From (3.35) and (3.36) it follows that

$$\lim_{m=\infty} \text{prob}\{n \leqq m \mid F, \zeta_0\} = 1 \tag{3.37}$$

uniformly in F. Because of (3.36) and (3.37), we have

$$r(F, \zeta_i) = \sum_{m=1}^{\infty} r_m(F, \zeta_i) \qquad (i = 0, 1, 2, \cdots, \text{ad inf.}), \tag{3.38}$$

where

$$\begin{aligned} r_m(F, \zeta_i) = &\sum_{l=1}^{k} \int_{M_m} p_m(x^1, \cdots, x^m \mid F) W(F, d_l)\, dH_{m,i}(S_m, d_l) \\ &+ \int_{M_m} p_m(x^1, \cdots, x^m \mid F) c(x^1, \cdots, x^m)\, dH_{m,i}(S_m, D). \end{aligned} \tag{3.39}$$

Since

$$\lim_{i=\infty} H_{m,i}(S_m, D^*) = H_{m,0}(S_m, D^*)$$

for any subset D^* of D, it follows from Lemma 3.2 that

$$\lim_{i=\infty} r_m(F, \zeta_i) = r_m(F, \zeta_0). \tag{3.40}$$

Lemma 3.3 is an immediate consequence of (3.38) and (3.40).

3.4. *Equality of* Sup Inf r *and* Inf Sup r, *and other theorems.* In this section we shall prove the main theorems for the continuous case, using the lemmas derived in the preceding section.

THEOREM 3.1. *If Conditions* 3.1–3.5 *are fulfilled, then*

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r(\xi, \zeta) = \operatorname*{Inf}_{\zeta} \operatorname*{Sup}_{\xi} r(\xi, \zeta). \tag{3.41}$$

PROOF: Let Z^m denote the class of all ζ's for which prob $\{n \leqq m \mid \zeta, F\} = 1$ for all F. We shall denote an element of Z^m by ζ^m. First we shall show that it

is sufficient if for any finite m we can prove Theorem 3.1 under the restriction that ζ must be an element of Z^m. For this purpose, put $W_0 = \operatorname{Sup}_{F,d} W(F, d)$ and choose a positive integer m_ϵ so that

$$c(x^1, \cdots, x^m) > \frac{W_0^2}{\epsilon} \tag{3.42}$$

for all $m \geqq m_\epsilon$. The existence of such a value m_ϵ follows from Condition 3.4. We shall now show that for any ξ we have

$$\operatorname{Inf}_{\zeta^m} r(\xi, \zeta^m) \leqq \operatorname{Inf}_{\zeta} r(\xi, \zeta) + \epsilon \text{ for any } m \geqq m_\epsilon. \tag{3.43}$$

Let ζ_1 be any decision rule. There are two cases to be considered: (a) prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} \geqq \frac{\epsilon}{W_0}$; (b) prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} < \frac{\epsilon}{W_0}$. In case (a) we have $r(\xi, \zeta_1) \geqq W_0$. In this case, let ζ_2 be the rule that we decide for some d without taking any observations. Clearly, we shall have $r(\xi, \zeta_2) \leqq W_0$ and, therefore, $r(\xi, \zeta_2) \leqq r(\xi, \zeta_1)$. In case (b), let ζ_2 be defined as follows: $h_r(x^1, \cdots, x^r, D^*)$ for ζ_2 is the same as that for ζ_1 when $r < m_\epsilon$, and $h_r(x^1, \cdots, x^r, d_0)$ for ζ_2 is equal to $1 - \sum_{k=1}^{m_\epsilon - 1} h_k(x^1, \cdots, x^k, D)$ when $r = m_\epsilon$, and zero when $r > m_\epsilon$ where d_0 is a fixed element of D. Since prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} < \frac{\epsilon}{W_0}$, we have

$$r(\xi, \zeta_2) \leqq r(\xi, \zeta_1) + \epsilon.$$

In both cases ζ_2 is an element of Z^{m_ϵ}. Hence (3.43) is proved. From (3.43) we obtain

$$\operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta} r \leqq \operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta^{m_\epsilon}} r \leqq \operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta} r + \epsilon. \tag{3.44}$$

Assume now that

$$\operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta^m} r = \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi} r \tag{3.45}$$

holds for any m. From (3.44) and (3.45) we obtain

$$\operatorname{Inf}_{\zeta^{m_\epsilon}} \operatorname{Sup}_{\xi} r \leqq \operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta} r + \epsilon. \tag{3.46}$$

Hence

$$\operatorname{Inf}_{\zeta} \operatorname{Sup}_{\xi} r \leqq \operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta} r + \epsilon. \tag{3.47}$$

Since this is true for any ϵ, we have

$$\operatorname{Inf}_{\zeta} \operatorname{Sup}_{\xi} r \leqq \operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta} r. \tag{3.48}$$

Theorem 3.1 follows from (3.48) and Lemma 1.3.

To complete the proof of Theorem 3.1, it remains to be shown that (3.45) holds for any m. Since D is compact, (3.45) is proved if we can prove it for any finite D. In the remainder of the proof we shall, therefore assume that D consists of k points $d_1, \cdots, d_k$. Let ω be a subset of Ω that is conditionally compact in the sense of the metric[7]

$$\delta_0(F_1, F_2) = \operatorname*{Sup}_{S_m} \left| \int_{S_m} dF_1 - \int_{S_m} dF_2 \right| \tag{3.49}$$

where S_m is a subset of the m-dimensional sample space. We shall show that ω is conditionally compact also in the sense of the intrinsic metric given by

$$\delta_1(F_1, F_2) = \operatorname*{Sup}_{\zeta^m} | r(F_1, \zeta^m) - r(F_2, \zeta^m) | . \tag{3.50}$$

Let

$$\delta_2(F_1, F_2) = \operatorname*{Sup}_{d} | W(F_1, d) - W(F_2, d) | \tag{3.51}$$

and

$$\delta_3(F_1, F_2) = \delta_0(F_1, F_2) + \delta_2(F_1, F_2). \tag{3.52}$$

It follows from Condition 3.3 and Theorem 3.1 in [3] that Ω, and therefore also ω, is conditionally compact in the sense of the metric $\delta_2(F_1, F_2)$. Hence ω is conditionally compact in the sense of the metric $\delta_3(F_1, F_2)$. The conditional compactness of ω relative to the metric $\delta_1(F_1, F_2)$ is proved, if we can show that any sequence $\{F_i\}$ that is a Cauchy sequence relative to the metric δ_3 is a Cauchy sequence also relative to the metric δ_1. Let $\{F_i\}$ ($i = 1, 2, \cdots$, ad inf.) be a Cauchy sequence relative to δ_3. Then there exists a distribution F_0 (not necessarily an element of Ω) and a function $W(d)$ such that

$$\lim_{i=\infty} W(F_i, d) = W(d) \text{ uniformly in } d \tag{3.53}$$

and

$$\lim_{i=\infty} \int_{S_m} dF_i = \int_{S_m} dF_0 \tag{3.54}$$

uniformly in S_m. We have

$$\begin{aligned} r(F_i, \zeta^m) = &\sum_{r=1}^{m} \sum_{j=1}^{k} \int_{M_r} \\ &\cdot p(x^1, \cdots, x^r \mid F_i) W(F_i, d_j) h_r(x^1, \cdots, x^r, d_j)\, dx^1 \cdots dx^r \\ &+ \sum_{r=1}^{m} \int_{M_r} c\,(x^1, \cdots, x^r) p(x^1, \cdots, x^r \mid F_i) h_r(x^1, \cdots, x^r, D)\, dx^1 \cdots dx^r, \end{aligned} \tag{3.55}$$

[7] By $\int_{S_m} dF$ we mean $\int_{S_m} p_m(x^1, \cdots, x^m \mid F) dx^1 \cdots dx^m$.

where M_r denotes the r-dimensional sample space. The sequence $\{F_i\}$ is a Cauchy sequence relative to the metric δ_1 if there exists a function $r(\zeta^m)$ such that

$$\lim_{i=\infty} r(F_i, \zeta^n) = r(\zeta^m) \tag{3.56}$$

uniformly in ζ^m. Let $\bar{r}(F_i, \zeta^m)$ be the function we obtain from $r(F_i, \zeta^m)$ by replacing the factor $W(F_i, d_j)$ by $W(d_j)$ under the first integral on the right hand side of (3.55). Because of (3.53), we have

$$\lim_{i=\infty} [r(F_i, \zeta^m) - \bar{r}(F_i, \zeta^m)] = 0 \tag{3.57}$$

uniformly in ζ^m. Thus, (3.56) is proved if we can show the existence of a function $\bar{r}(\zeta^m)$ such that

$$\lim_{i=\infty} \bar{r}(F_i, \zeta^m) = \bar{r}(\zeta^m) \tag{3.58}$$

uniformly in ζ^m. Let C be a class of functions $\varphi(x^1, \cdots, x^m)$ such that

$$|\varphi(x^1, \cdots, x^m)| < A < \infty \text{ for all } \varphi \text{ in } C.$$

It then follows from (3.54) that there exists a functional $g(\varphi)$ such that

$$\lim_{i} \int_{M_m} \varphi \, dF_i = g(\varphi) \tag{3.59}$$

uniformly in φ. Application of this general result yields (3.58) immediately. Hence, $\{F_i\}$ is a Cauchy sequence relative to the metric δ_1 and, therefore ω is shown to be conditionally compact relative to the metric δ_1 if it is relative to the metric δ_0.

It then follows from Theorem 3.2 in [3] that $\operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta^m} r = \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi} r$ if we replace Ω by a subset ω that is conditionally compact relative to δ_0.[8] Since Ω is separable relative to δ_0, there exists a sequence $\{\Omega_i\}$ of subsets of Ω such that Ω_i is conditionally compact relative to δ_0, $\Omega_{i+1} \supset \Omega_i$ and $\sum_{i=\infty} \Omega_i = \Omega^*$ is dense in Ω. Let ξ' denote an a priori distribution ξ for which $\xi(\Omega_i) = 1$. Since the left and right hand members in (3.45) remain unchanged when Ω is replaced by Ω^*, it follows from Theorem 1.3 that equation (3.45) is proved if we can show that

$$\lim_{i=\infty} \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi'} r = \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi} r. \tag{3.60}$$

Let $\{\zeta_i^m\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision rules such that

$$\lim_{i=\infty} [\operatorname{Sup}_{\xi'} r(\xi^i, \zeta_i^m) - \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi'} r] = 0. \tag{3.61}$$

---

[8] Strictly, we would have to write $\operatorname{Inf}_{\eta^m}$ instead of $\operatorname{Inf}_{\zeta^m}$ where η^m is a probability measure in the space of all ζ^m. But, since the use of any discrete probability measure is equivalent to the use of a ζ^m, and since the restriction to discrete η^m does not change $\operatorname{Sup}_{\xi} \operatorname{Inf}_{\eta^m} r$ or $\operatorname{Inf}_{\eta^m} \operatorname{Sup}_{\xi} r$ we can replace $\operatorname{Inf}_{\eta^m}$ by $\operatorname{Inf}_{\zeta^m}$.

According to Lemmas 3.1 and 3.3, there exists a subsequence $\{i_j\}$ of $\{i\}$ and a decision rule ζ_0^m such that

$$\liminf_{j=\infty} r(F, \zeta_{i_j}^m) \geqq r(F, \zeta_0^m) \text{ for all } F. \tag{3 62}$$

Since Ω^* is dense in Ω, it follows from (3.61) and (3.62) that

$$\sup_F r(F, \zeta_0^m) \leqq \lim_{i=\infty} \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi^i} r \tag{3.63}$$

and, therefore, 3.60 holds. Thus, (3.45) is proved and the proof of Theorem 3.1 is completed.

THEOREM 3.2. *If Conditions* 3.1–3.5 *are fulfilled, then there exists a minimax solution, i e., a decision rule* ζ_0 *for which*

$$\sup_F r(F, \zeta_0) \leqq \sup_F r(F, \zeta) \text{ for all } \zeta. \tag{3.64}$$

PROOF: Because of Theorem 3.1 there exists a sequence $\{\zeta_i\}$ ($i = 1, 2, \cdots$, ad inf.) of decision rules such that

$$\lim_{i=\infty} \sup_F r(F, \zeta_i) = \inf_\zeta \sup_F r(F, \zeta). \tag{3.65}$$

According to Lemmas 3.1 and 3.3 there exists a subsequence $\{\zeta_{i_j}\}$ of $\{\zeta_i\}$ and a decision rule ζ_0 such that

$$\liminf_{j=\infty} r(F, \zeta_{i_j}) \geqq r(F, \zeta_0) \text{ for all } F. \tag{3.66}$$

It follows from (3.65) and (3.66) that ζ_0 is a minimax solution and Theorem 3.2 is proved.

THEOREM 3.3. *If Conditions* 3.1–3.5 *are fulfilled, then for any* ξ *there exists a Bayes solution relative to* ξ.

This theorem is an immediate consequence of Lemmas 3.1 and 3 3.

THEOREM 3.4. *If Conditions* 3.1–3.5 *are fulfilled, then the class of all Bayes solutions in the wide sense is a complete class.*

The proof is omitted, since it is entirely analogous to that of Theorem 2.5.

3.5. *Formulation of an additional condition.* In this section we shall formulate an additional condition which will permit the derivation of some stronger theorems. Let the metric $\delta_0(F_1, F_2)$ be defined by

$$\delta_0(F_1, F_2) = \sum_{m=1}^{\infty} \frac{1}{m^2} \operatorname{Sup}_{S_m} \left| \int_{S_m} d F_1 - \int_{S_m} d F_2 \right|$$

where S_m may be any subset of the m-dimensional sample space.

CONDITION 3.6. *The space* Ω *is compact relative to the metric* $\delta_0(F_1, F_2)$

$$\lim_i W(F_i, d) = W(F_0, d)$$

uniformly in d *if* $\lim_i \delta_0(F_i, F_0) = 0$.

THEOREM 3.5. *If Conditions* 3.1–3.6 *hold, then*

(i) *there exists a least favorable a priori distribution*

(ii) *any minimax solution is a Bayes solution in the strict sense*

(iii) *for any decision rule ζ_0 which is not a Bayes solution in the strict sense and for which $r(F, \zeta_0)$ is a bounded function of F there exists a decision rule ζ_1 which is a Bayes solution in the strict sense and is uniformly better than ζ_0.*

PROOF: The proofs of (i) and (ii) are entirely analogous to those of (i) and (ii) in Theorem 2.6, and will therefore be omitted here.

To prove (iii), let ζ_0 be a decision rule that is not a Bayes solution in the strict sense and for which $r(F, \zeta_0)$ is bounded. We replace the weight function $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \zeta_0)$. We shall show that (i) remains valid when $W(F, d)$ is replaced by $W^*(F, d)$ This is not obvious, since $r(F, \zeta_0)$, and therefore also $W^*(F, d)$ may not be continuous in F. First we shall prove that

$$\liminf_{i=\infty} r(\xi_i', \zeta_0) \geqq r(\xi_0', \zeta_0) \tag{3.67}$$

for any sequence $\{\xi_i'\}$ for which

$$\lim_{i=\infty} \xi_i'(\omega) = \xi_0'(\omega)$$

for any open subset ω of Ω (in the sense of the metric δ_0) whose boundary has probability measure zero according to ξ_0'. Let $r_m(F, \zeta)$ denote the conditional expected value of the loss $W(F, d)$ plus the cost of experimentation when $n = m$, F is true and the rule ζ is used by the statistician (see equation (3.39)). Since $W(F, d)$ and the cost of experimentation when m observations are taken are uniformly bounded, one can easily verify that

$$\lim_{i=\infty} r_m(F_i, \zeta_0) = r_m(F_0, \zeta_0) \tag{3.68}$$

for any sequence $\{F_i\}$ for which

$$\lim_{i=\infty} \delta_0(F_i, F_0) = 0. \tag{3.69}$$

Hence, since Ω is compact (Condition 3.6),

$$\lim_{i=\infty} r_m(\xi_i', \zeta_0) = r_m(\xi_0', \zeta_0) \tag{3.70}$$

where

$$r_m(\xi, \zeta_0) = \int_\Omega r_m(F, \zeta_0)\, d\xi. \tag{3.71}$$

Since

$$r(\xi, \zeta_0) = \sum_{m=1}^{\infty} r_m(\xi, \zeta_0)$$

inequality (3.67) follows from (3.70).

The remainder of the proof of (iii) will be omitted here, since it is the same as that of (iii) in Theorem 2.6.

We shall now replace Condition 3.6 by the following weaker one.

CONDITION 3.6*. *There exists a sequence* $\{\Omega_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *of subsets of* Ω *such that Condition* 3.6 *is fulfilled when* Ω *is replaced by* Ω_i, $\Omega_{i+1} \supset \Omega_i$, *and* $\lim_{i=\infty} \Omega_i = \Omega$.

THEOREM 3.6. *If Conditions* 3.1–3.5 *and* 3.6* *are fulfilled then*

(i) *A minimax solution* ζ_0 *and a sequence* $\{\zeta_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *exist such that* $\lim_{i=\infty} \zeta_i = \zeta_0$ *and* ζ_i $(i = 1, 2, \cdots, ad\ inf.)$ *is a Bayes solution in the strict sense.*

(ii) *For any decision rule* ζ_0 *for which* $r(F, \zeta_0)$ *is bounded there exists another decision rule* ζ_1 *such that* ζ_1 *is a limit of a sequence of Bayes solutions in the strict sense and* $r(F, \zeta_1) \leqq r(F, \zeta_0)$ *for all* F *in* Ω.

PROOF: According to Theorem 3.5, for each i there exists a decision rule ζ_i $(i = 1, 2, \cdots, \text{ad inf.})$ such that ζ_i is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i. Let $\{\zeta_{i_j}\}$ be a subsequence of the sequence $\{\zeta_i\}$ such that $\{\zeta_{i_j}\}$ admits a limit ζ_0, i.e., $\lim_{j=\infty} \zeta_{i_j} = \zeta_0$. Because of Lemma 3.3,

$$\liminf_{j=\infty} r(F, \zeta_{i_j}) \geqq r(F, \zeta_0). \tag{3.72}$$

Hence ζ_0 is a minimax solution relative to the original space Ω and statement (i) is proved.

To prove (ii), replace $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \zeta_0)$ where ζ_0 is a decision rule for which $r(F, \zeta_0)$ is bounded. In proving statement (iii) of Theorem 3.5, we have shown that there exists a decision rule ζ_{1i} $(i = 1, 2, \cdots, \text{ad inf.})$ such that ζ_{1i} is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i and $W(F, d)$ by $W^*(F, d)$. Clearly, ζ_{1i} remains to be a Bayes solution in the strict sense also relative to Ω and $W(F, d)$. Since ζ_{1i} is a minimax solution relative to Ω_i and $W^*(F, d)$, we have

$$r(F, \zeta_{1i}) \leqq r(F, \zeta_0) \text{ for all } F \text{ in } \Omega_i. \tag{3.73}$$

Let $\{\zeta_{1i_j}\}$ be a convergent subsequence of $\{\zeta_{1i}\}$ and let $\lim_{j=\infty} \zeta_{1i_j} = \zeta_1$. Then, because of Lemma 3.3, we have

$$r(F, \zeta_1) \leqq r(F, \zeta_0) \text{ for all } F \text{ in } \Omega.$$

Since ζ_1 is a limit of a sequence of Bayes solutions in the strict sense, statement (ii) is proved.

Addition at proof reading. After this paper was sent to the printer the author found that Ω is always separable (in the sense of the convergence definition in Condition 3.5) and, therefore, Condition 3.5 is unnecessary. A proof of the separability of Ω will appear in a forthcoming publication of the author.

The boundedness of $r(F, \zeta_i)$ is not necessary for the validity of Lemma 3.3. Let $\lim_{i=\infty} \zeta_i = \zeta_0$ and suppose that for some F, say F_0, $r(F_0, \zeta_i)$ is not bounded in i. If $\liminf_{i=\infty} r(F_0, \zeta_i) = \infty$, Lemma 3.3 obviously holds for $F = F_0$. If

$\lim_{i=\infty}\inf r(F_0, \zeta_i) = g < \infty$, let $\{i_j\}$ be a subsequence of $\{i\}$ such that $\lim_{j=\infty} r(F_0, \zeta_{i_j}) = g$. Since $r(F_0, \zeta_{i_j})$ is a bounded function of j, Lemma 3.3 is applicable and we obtain $g \geqq r(F_0, \zeta_0)$. In a similar way, one can see that also Lemma 2.4 remains valid without assuming the boundedness of $r(F, \eta_i)$.

Although not stated explicitly, several functions considered in this paper are assumed to be measurable with respect to certain additive classes of subsets. In the continuous case, for example, the precise measurability assumptions may be stated as follows: Let B be the class of all Borel subsets of the infinite dimensional sample space M. Let H be the smallest additive class of subsets of Ω which contains any subset of Ω which is open in the sense of at least one of the convergence definitions considered in this paper. Let T be the smallest additive class of subsets of D which contains all open subsets of D (in the sense of the metric $\delta(d_1, d_2, \Omega)$). By the symbolic product $H \times T$ we mean the smallest additive class of subsets of the Cartesian product $\Omega \times D$ which contains the Cartesian product of any member of H by any member of T. The symbolic product $H \times B$ is similarly defined. It is assumed that: (1) $W(F, d)$ is measurable $(H \times T)$; (2) $p_m(x^1, \cdots, x^m \mid F)$ is measurable $(B \times H)$; (3) $\delta_{x^1 \cdots x^r}(D^*)$ is measurable (B) for any member D^* of T; (4) $z_r(x^1, \cdots, x^r)$ and $c_r(x^1, \cdots, x^r)$ are measurable (B). These assumptions are sufficient to insure the measurability (H) of $r(F, \zeta)$ for any ζ.

REFERENCES

[1] J. v. Neumann and Oskar Morganstein, *Theory of Games and Economic Behavior*, Princeton University Press, 1944

[2] A. Wald, "Generalizatoin of a theorem by v. Neumann concerning zero sum two-person games," *Annals of Mathematics*, Vol. 46 (April, 1945).

[3] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (October, 1947).

THE MULTIPLICATIVE PROCESS

By Richard Otter[1, 2]

University of Notre Dame

1. Introduction and summary. The multiplicative process is usually defined by the sequence of random variables $X_0, X_1, \cdots$ whose distributions are specified as follows: $P(X_0 = 1) = 1$, $\sum_{\nu=0}^{\infty} P(X_1 = \nu) = 1$, and if $X_n = 0$ then $P(X_{n+1} = 0) = 1$, whereas if X_n is a positive integer then X_{n+1} is distributed as the sum of X_n independent random variables each with the distribution of X_1. The variable X_n is interpreted as the number of "particles" in the nth generation, and the index n as a discrete time parameter. This has been the method of approach in previous studies of the process [1, 2, 3, 4, 5] The multiplicative process has various applications, notably in the study of population growth, the spread of epidemics or rumors, and the nuclear chain reaction. The closely related "birth and death" process was recently studied by Kendall [6].

Whenever one studies the probability theory of a particular system there seem to be definite conceptual advantages in defining explicitly the set $\mathcal{T}$ of elementary events, the additive class $\mathfrak{M}$ of subsets of $\mathcal{T}$, called events, and the probability measure P for the events of $\mathfrak{M}$. Now an elementary event of this process can be represented by a rooted tree where the original particle is represented by the root vertex and where the particles of the nth generation are represented by the vertices n segments removed from the root. The tree will be finite or infinite according to whether a finite or an infinite number of particles are involved in the elementary event. Thus, the set of trees is the natural choice for $\mathcal{T}$. The first part of this paper is devoted to a more precise description of $\mathcal{T}$, $\mathfrak{M}$ and P. We shall then see easily that $X_n(t)$, the number of vertices n segments removed from the root of $t \in \mathcal{T}$, i.e. the number of particles in the nth generation, has the distribution defined in the preceding paragraph. Since the time does not appear in our description of $\mathcal{T}$ we fetter ourselves somewhat if we interpret n as a discrete time parameter. Thus, we have already reaped some harvest from considering the process from the point of view of $\mathcal{T}$. Another advantage is that we are led in a natural way to study the distribution of other structural features of the trees, e.g the total number of vertices, or the number of vertices with k outgoing segments.

The chief results of this paper are as follows. The recursion formula for the probability P_n that a tree have n vertices $n = 1, 2, \cdots$ is obtained as well as an asymptotic estimate of P_n valid for large n. The distributions of the number of branches at the root in a finite tree, an infinite tree, or in a tree with n vertices are obtained and the asymptotic distribution of the latter as $n \to \infty$. The

[1] Research under an Office of Naval Research contract.

[2] The author wishes to express his gratitude to Professor E. Artin of Princeton University for the suggestion of this problem and his encouragement towards its solution.

distribution of the fraction of vertices with k outgoing segments in the finite trees, in the trees with n vertices, and the asymptotic distribution of the latter as $n \to \infty$ are also found. Finally, an estimate is obtained for the probability that a tree be finite in case this probability is near 1, a result which was previously obtained by Kolmogoroff [7].

2. The space of trees. We shall use the notation $\{a\}$, $\{a_1, a_2, \cdots a_n\}$, $\{a_j\}_{j\in J}$, and $\{a_j \mid R\}_{j\in J}$ to denote the sets which consist of respectively the single element a, the elements $a_1, a_2, \cdots a_n$, all a_j with $j \in J$, and all a_j with the property R and $j \in J$. We denote the union of two sets A and B by $A + B$, their intersection by AB, and the cartesian product of n identical factors each of which is A by $A^{(n)}$.

Let I denote the set of positive integers. We assume given for each $n \in I$ a countable set U_n of objects $u_{i_1 i_2 \cdots i_n}$ called vertices, i.e.

$$U_n = \{u_{i_1 i_2 \cdots i_n}\}_{(i_1, i_2 \cdots i_n) \in I^{(n)}}.$$

Let u_0 be a vertex distinct from all the other vertices and let $U = \{u_0\} + \Sigma U_n$ be the collection of all the vertices. We shall interpret u_0 as the original parent particle and the vertex u_{152}, for example, as the second son of the fifth son of the first son of the original particle. If s is a subset of U, $s \subset U$, and if $i_1, i_2, \cdots i_{n+m}$ are such that $u_{i_1 i_2 \cdots i_n}$, $u_{i_1 i_2 \cdots i_n i_{n+1}}$, $\cdots$ $u_{i_1 i_2 \cdots i_n i_{n+1} \cdots i_{n+m}}$ each belong to s then this set of vertices is called a *path* from $u_{i_1 i_2 \cdots i_n}$ to $u_{i_1 i_2 \cdots i_{n+m}}$ in s and $m \geq 0$ is the *length* of the path or the *distance* from $u_{i_1 i_2 \cdots i_n}$ to $u_{i_1 i_2 \cdots i_{n+m}}$. If $m = 1$ we call the path a *segment*, for short.

For the sake of convenience let us agree to put $u_{i_1 i_2 \cdots i_n 0} = u_{i_1 i_2 \cdots i_n}$, $(n \geq 1)$ then we define $W(s, u)$, for $u \in s \subset U$, to be the number of segments from u in s, and we call $W(s, u)$ the *type* of the vertex u in s. If t is a subset of U, then we call t a *tree* if and only if

$$W(t, u) < \infty \quad \text{for} \quad u \in t \tag{1}$$

and

$$u_{i_1 i_2 \cdots i_n} \in t \quad \text{implies} \quad u_{i_1 i_2 \cdots i_{n-1} \nu} \in t \quad \text{for} \quad \nu = 0, 1, \cdots i_n. \tag{2}$$

Let $\mathcal{T}$ be the set of all trees. The condition (2) clearly implies that for each $t \in \mathcal{T}$ we have $u_0 \in t$ and that there is a unique path from u_0 to any other vertex of t. Hence, whenever a path exists between any two vertices of t it is unique. We call u_0 the *root* of t. If for $u \in t \in \mathcal{T}$ we have $W(t, u) = 0$ then u is called an *endpoint* of t, and the vertices of t which are not endpoints are called *inner* vertices. (It is to be noted that the objects we call trees here are rooted trees in the sense of Cayley but our trees have their vertices numbered as well. Usually one would identify the trees $\{u_0, u_1, u_2, u_{11}\}$ and $\{u_0, u_1, u_2, u_{21}\}$, but we do not wish to do so because for us it is distinctly different whether the grandson is sired by the first son or by the second son.)

For $u \in t \in \mathcal{T}$ we define the *branch* of t at u to be the set of all vertices belonging

to any path from u in t. Our convention of admitting paths of length 0 implies that $u \in b(t, u)$. In fact, if $W(t, u) = 0$ then $b(t, u) = \{u\}$. If t' is a tree such that $t' \subset t$ then we call t an *extension* of t', denoted $t \geq t'$ or $t' \leq t$, if $W(t', u) > 0$ implies $W(t', u) = W(t, u)$. Thus $t \geq t'$ is equivalent to $t \supset t'$ and

$$t = t' + \sum_{u} b(t, u)$$

where u runs through all the endpoints of t'. The extension relation imposes a partial ordering upon $\mathcal{T}$.

The extension t of t' is interpreted as a possible future aspect of a family tree when its structure at present is given by t', all present members of the family who have progeny being regarded as sterile.

If $u = u_{i_1, i_2 \cdots i_n}$ then the mapping φ defined for the vertices of $b(t, u)$ by putting

$$\varphi(u_{i_1 i_2 \cdots i_n i_{n+1} \cdots i_{n+m}}) = u_{i_{n+1} \cdots i_{n+m}}$$

maps $b(t, u)$ one to one onto a tree $\varphi(b(t, u))$ in such a fashion that if $\{v_1, v_2\}$ is a segment from v_1 to v_2 in $b(t, u)$ then $\{\varphi(v_1), \varphi(v_2)\}$ is a segment from $\varphi(v_1)$ to $\varphi(v_2)$ in $\varphi(b(t, u))$. We call the mapping φ a *homeomorphism* and we say that $b(t, u)$ is *homeomorphic* to $\varphi(b(t, u))$.

If a tree contains a finite number of vertices then it is called a *finite* tree; otherwise it is an *infinite* tree. Let $\mathcal{F}$ denote the set of all finite trees and $\mathcal{I}$ the set of all infinite trees, and let $\mathcal{K}$ denote the set of non-negative integers. For each $k \in \mathcal{K}$ we define $Y_k(t)$ for $t \in \mathcal{F}$ to be the number of vertices of type k in t. When it is clear to which tree t we refer we shall usually abbreviate $Y_0(t)$ by m, and we agree not to use the letter m with any other connotation. For each $T \in \mathcal{F}$ let $e_1(T), e_2(T), \cdots e_m(T)$ denote its m endpoints. We then define for $T \in \mathcal{F}$ and $\kappa = (k_1, k_2, \cdots k_m) \in \mathcal{K}^{(m)}$

$$[T, \kappa] = \{t \mid t \geq T, W(t, e_i(T)) = k_i, i = 1, 2, \cdots m' \in \mathcal{T}\},$$

and we call $[T, \kappa]$ a *neighborhood*. For each $t \in [T, \kappa]$ we say $[T, \kappa]$ is a *neighborhood of* t. Then it is easy to show that $\mathcal{T}$ is a *topological space* where the neighborhoods defined above form the *defining system of neighborhoods* [8].

3. The measure theory in $\mathcal{T}$. In the following paragraphs an outline of the measure theory in $\mathcal{T}$ is given which omits proofs for the most part since they are easily constructed. The only point of difficulty arises in showing the measure function to be completely additive, but here the outline has more detail.

Let $\mathfrak{S}$ be the collection of subsets of $\mathcal{T}$ such that $0 \in \mathfrak{S}$ and any other set S belongs to $\mathfrak{S}$ if and only if there is a $t \in \mathcal{F}$ and a non-void "rectangle set" $A = A_1 \times A_2 \times \cdots A_m \subset \mathcal{K}^{(m)}$, $m = Y_0(t)$, such that

$$S = \sum_{\kappa \in A} [t, \kappa] \tag{3}$$

where the sets $A_1, A_2, \cdots A_m$ may be finite or infinite sets of non-negative

integers. The collection of neighborhoods which appear as terms in (3), i.e. $\{[t, \kappa]\}_{\kappa \in \Delta}$, we call an $\mathfrak{S}$-*partition* of S, and t is called the *generator* of the $\mathfrak{S}$-partition. Only a finite number of $\mathfrak{S}$-partitions are possible for an $S \in \mathfrak{S}$, because only a finite number of trees can possibly be generators and there is only one $\mathfrak{S}$-partition per generator. With respect to our partial ordering of the trees all possible generators lie between two particular ones. We call the smaller of these the *irreducible generator* and the corresponding $\mathfrak{S}$-partition the *irreducible* $\mathfrak{S}$-*partition* of S. Any partition of S into neighborhoods must be a subpartition of this irreducible $\mathfrak{S}$-partition The elements of $\mathfrak{S}$ also display two important properties of the rectangles in Euclidean space, namely if S $S' \in \mathfrak{S}$ then

$$SS' \in \mathfrak{S} \tag{4}$$

and if $S \subset S'$ then there is a finite chain

$$S = S_0 \subset S_1 \subset \cdots S_n = S' \tag{5}$$

such that S_i, $S_i - S_{i-1} \in \mathfrak{S}$ for $i = 1, 2, \cdots n$.
A class of sets with the properties (4) and (5) has been called a *half-ring* by von Neumann [9].

Let $p_0, p_1, \cdots$ be given non-negative numbers such that $\sum_0^\infty p_\nu = 1$. For $t \in \mathcal{T}$ let us put

$$\bar{P}(t) = \prod_{\nu=1}^{\infty} p_\nu^{\gamma_\nu(t)} \tag{6}$$

with the convention $0^0 = 1$. We then define the measure function P for the sets in $\mathfrak{S}$ by

$$\begin{aligned} &P(0) = 0 \\ &P([t, \kappa]) = \left(\prod_{\nu=1}^{m} p_{\kappa_\nu}\right) \bar{P}(t), \quad \text{where } \kappa = (k_1, k_2 \cdots k_m) \in \mathcal{K}^{(m)} \\ &P(S) = \sum_{\kappa \in \Delta} P([t, \kappa]), \quad \text{where } \{[t, \kappa]\}_{\kappa \in \Delta} \text{ is the irreducible } \mathfrak{S}\text{-partition of } S. \end{aligned} \tag{7}$$

P is evidently non-negative. Letting t be the tree with one vertex and putting $A = \mathcal{K}$ gives $P(\mathcal{T}) = 1$. It is easy to see that P is completely additive for the $\mathfrak{S}$-partitions of a neighborhood, but this implies P is completely additive for the $\mathfrak{S}$-partitions of an arbitrary element S of $\mathfrak{S}$. In order to show that P is completely additive for any partition of S into elements of $\mathfrak{S}$, it is necessary and sufficient to show this for an arbitrary partition of a neighborhood into neighborhoods. One may reach finer and finer partitions of a given neighborhood N by replacing a neighborhood in any one partition by an $\mathfrak{S}$-partition of the neighborhood, and repeating the process. The sum of the measures of the sets in the partition is invariant under such a replacement. On the other hand it can be shown that all possible partitions of N into neighborhoods may be reached in this way. More precisely, let $\bar{N} = \{N_i\}_{i \in J}$ be a partition of a neighborhood N

into neighborhoods N_i. We call $\bar{N}$ reduced if whenever a subset of $\bar{N}$ is an $\mathfrak{S}$-partition of a neighborhood $M \subset N$ then the partition consists of M itself, i.e. it is the irreducible $\mathfrak{S}$-partition of M. Then we have the following theorem:

THEOREM 1 *If $\bar{N}$ is a reduced partition of a neighborhood N into neighborhoods then $\bar{N} = \{N\}$.*

The proof is indirect and proceeds by constructing a decreasing sequence of neighborhoods contained in N whose limit is not void and yet has nothing in common with any N_i, but this is a contradiction.

Let $\mathfrak{F}$ consist of all those sets which may be formed by finite unions of disjoint elements of a half-ring $\mathfrak{S}$, then $\mathfrak{F}$ is a field of sets. If P is a completely additive measure on $\mathfrak{S}$ then its natural extension P_1 is completely additive on $\mathfrak{F}$ [9]. Kolmogoroff [10] has shown that the completely additive measure P_1 may always be extended to a completely additive measure P_2 on the Borel field $\mathfrak{M}$, i.e. the smallest additive class of sets containing $\mathfrak{F}$. Since $P_2(\mathcal{T}) = 1$, P_2 is a probability measure. For simplicity we put $P_2 = P$. Let us also agree that if M is the set of all trees with the property R we may write $P(R)$ instead of $P(M)$. If N is a set with $P(N) > 0$ then $P(M/N)$ shall denote the conditional probability of M, given N, i.e. $P(M/N) = (P(N))^{-1}P(MN)$.

4. Independence of the branches. In the multiplicative process the events occurring in one branch of a tree are independent of those in a second branch disjoint with the first and it is for this reason that the process is relatively simple to analyze. In this section we shall try to expose the character of this independence.

For $T \in \mathcal{F}$, let $\mathfrak{E}_T$ be the set of all extensions of T, then

$$\mathfrak{E}_T = \sum_{\kappa \in \mathcal{K}^{(m)}} [T, \kappa],$$

whence by (6) and (7) $P(\mathfrak{E}_T) = \bar{P}(T)$. The following lemma is then easily established.

LEMMA 1. *If $P(\mathfrak{E}_T) > 0$ then $W(t, e_i(T))$, $i = 1, 2, \cdots m$, under the condition $t \in \mathfrak{E}_T$, are independent-random variables each with the distribution,*

$$P(W(t, e_i(T)) = k/\mathfrak{E}_T) = p_k \qquad k = 0, 1, 2, \cdots . \tag{8}$$

In the particular case where $T = \{u_0\}$ we have $\mathfrak{E}_T = \mathcal{T}$ and we put $W(t) = W(t, u_0)$ for short. Thus $W(t)$ tells what type of vertex the root of t is and (8) becomes

$$P(W = k) = p_k \qquad k = 0, 1, 2, \cdots ,$$

For $t \in \mathcal{T}$ and $n = 0, 1, 2, \cdots$ let $X_n(t)$ be the number of vertices of t at distance n from its root. Then $X_0(t) = 1$ and $X_1(t) = W(t)$. If n, r are positive integers then there is at least one $T \in \mathcal{F}$ which has r of its endpoints, say $e_{i_1}(T), e_{i_2}(T), \cdots e_{i_r}(T)$, at distance n from the root and which also satisfies $X_{n+1}(T) = 0$. Put

$$\mathfrak{E}_T^{i_1 \cdots i_r} = \{t \mid W(t, e_j(T)) = 0, j \neq i_1, i_2, \cdots i_r, t \in \mathfrak{E}_T\}.$$

Evidently for $t \in \mathfrak{S}_T^{i_1 \cdots i_r}$

$$X_{n+1}(t) = \sum_{\nu=1}^{i} W(t, e_{i_\nu}(T)),$$

and a proof similar to that of lemma 1 gives

LEMMA 2. *If* $P(\mathfrak{S}_T^{i_1 \cdots i_r}) > 0$ *then* $X_{n+1}(t)$, *under the condition* $t \in \mathfrak{S}_T^{i_1 \cdots i_r}$, *is the sum of* r *independent random variables each with the distribution of* X_1.

By (6) and (7) for $t \in \mathfrak{M} \subset \mathcal{F}\mathfrak{S}_T$

$$P(t) = \prod_{\nu=0}^{\infty} p_\nu^{Y_\nu(t)},$$

which depends only upon the type of each vertex as it occurs in t. For those vertices which are inner vertices of T, $Y_\nu(t)$ is constant. Any other vertex belongs to one and only one of $b(t, e_1(T))$, $b(t, e_2(T))$, $\cdots$ $b(t, e_m(T))$ and its type in t is, of course, the same as its type in the branch to which it belongs. Furthermore, each branch is homeomorphic to just one tree in $\mathcal{F}$,

$$b(t, e_i(T)) \leftrightarrow t_i, \qquad i = 1, 2, \cdots m.$$

Since the type of a vertex is preserved under homeomorphism we have

$$P(t) = P(\mathfrak{S}_T)P(t_1)P(t_2) \cdots P(t_m).$$

If, as t runs through $\mathfrak{M}$, $(t_1, t_2, \cdots t_m)$ runs through $\mathfrak{M}_1 \times \mathfrak{M}_2 \times \cdots \mathfrak{M}_m$, we obtain

$$P(\mathfrak{M}) = P(\mathfrak{S}_T)P(\mathfrak{M}_1)P(\mathfrak{M}_2) \cdots P(\mathfrak{M}_m). \tag{10}$$

Let us hereafter put $p = P(\mathcal{F})$. In the particular case of (10) where $\mathfrak{M} = \mathcal{F}\mathfrak{S}_T$ we clearly have $\mathfrak{M}_i = \mathcal{F}$, $i = 1, 2, \cdots m$, hence

$$P(\mathcal{F}\mathfrak{S}_T) = P(\mathfrak{S}_T) \cdot p^m. \tag{11}$$

If we define T_ν, $\nu = 0, 1, 2, \cdots$, to be the tree with $\nu + 1$ vertices which has $W(T_\nu) = \nu$ then

$$\begin{aligned} \mathcal{T} &= \{u_0\} + \sum_{\nu=1}^{\infty} \mathfrak{S}_{T_\nu}, \\ \mathcal{F} &= \{u_0\} + \sum_{\nu=1}^{\infty} \mathcal{F}\mathfrak{S}_{T_\nu}, \end{aligned} \tag{12}$$

where

$$\begin{aligned} \mathfrak{S}_{T_i}\mathfrak{S}_{T_j} = \mathfrak{S}_{T_j}\{u_0\} &= 0, \qquad i \neq j; \\ P(\mathfrak{S}_{T_\nu}) &= p_\nu, \qquad \nu = 1, 2, \cdots. \end{aligned}$$

From (11) and (12) we get

$$\sum_{\nu=0}^{\infty} p_\nu p^\nu = p. \tag{13}$$

For $t \in \mathfrak{FE}_T$ let $Z(b(t, e_i(T)))$ be the number of vertices in the branch of t at $e_i(T)$. In the particular case where $T = \{u_0\}$ we have $b(t, u_0) = t$ and $Z(t)$ is the number of vertices of t If now

$$\mathcal{F}_n = \{t \mid Z(t) = n, t \in \mathcal{F}\}; \qquad n = 1, 2, \cdots;$$
$$P_n = P(\mathcal{F}_n),$$

then by putting $\mathfrak{M} = \mathfrak{M}_T^{n_1 \cdots n_m}$ where

$$\mathfrak{M}_T^{n_1 \cdots n_m} = \{t \mid Z(b(t, e_i(T))) = n_i, \qquad i = 1, 2, \cdots m, t \in \mathfrak{FE}_T\}, \tag{14}$$
$$\mathfrak{M}_i = \mathcal{F}_{n_i}, \qquad i = 1, 2, \cdots m,$$

we may apply (10), which gives

$$P(t \in \mathfrak{FE}_T, Z(b(t, e_i(T))) = n_i, i = 1, 2, \cdots m) = P(\mathfrak{E}_T)P_{n_1}P_{n_2} \cdots P_{n_m}. \tag{15}$$

If $p > 0$ we may multiply and divide the right hand member of (15) by p^m which leads us to the following lemma:

LEMMA 3. *If $P(\mathfrak{FE}_T) > 0$, then $Z(b(t, e_i(T)))$, $i = 1, 2, \cdots m$, under the condition $t \in \mathfrak{FE}_T$, are m independent random variables each with the distribution of $Z(t)$, given $t \in \mathcal{F}$.*

5. The distribution of $Z(t)$. Let $f(w)$ be the generating function for the distribution of W, i.e.

$$f(w) = \sum_{\nu=0}^{\infty} p_\nu w^\nu \tag{16}$$

where w is a complex variable. If one is interested in studying the sequence $X_0, X_1, \cdots$ then one should define another sequence of functions $f_0, f_1, \cdots$ where $f_0(w) = w$ and $f_{n+1}(w) = f(f_n(w))$ for $n = 0, 1, 2, \cdots$. By computing formally the expansion of $f_n(w)$ around $w = 0$ it is not difficult to show that $f_n(w)$ is the generating function for X_n, i.e. $f_n(w) = \sum_{\nu=0}^{\infty} P(X_n = \nu)w^\nu$ which is the starting point for the previous investigations of the multiplicative process. But since we shall be mainly interested in the distribution of Z we define $\mathcal{P}(z)$ to be the corresponding generating function, i.e.

$$\mathcal{P}(z) = \sum_{n=1}^{\infty} P_n z^n. \tag{17}$$

Let ρ and α be the radii of convergence of the power series in the right members of (16) and (17) respectively. Since $f(1) = 1$ and $\mathcal{P}(1) = P(\mathcal{F}) \leq 1$ we know $\rho, \alpha \geq 1$ hence $f(w)$ and $\mathcal{P}(z)$ are analytic in $|w| < \rho$ and $|z| < \alpha$ respectively. The relation between the distribution of W and that of Z is put in evidence by the following theorem:

THEOREM 2. *Let*

$$\mathfrak{F}(z, w) = zf(w) - w,$$

then $w = \mathcal{P}(z)$ *is the unique analytic solution of*

$$\mathcal{G}(z, w) = 0 \tag{18}$$

in a certain neighborhood of (0, 0).

PROOF. Since $\mathcal{P}(z)$ is analytic at 0 and $\mathcal{P}(0) = 0$ it suffices to show that if we substitute formally $\sum P_n Z^n$ for w in $z\sum p_\nu w^\nu$ the coefficient of z^n is uniquely determined and is P_n.

$$z\sum_{\nu=0}^{\infty} p_\nu(\mathcal{P}(z))^\nu = p_0 z + \sum_{n=2}^{\infty}\left(\sum_{\nu=1}^{\infty}\sum_{\Sigma n_i = n-1} p_\nu P_{n_1}P_{n_2}\cdots P_{n_\nu}\right)z^n. \tag{19}$$

If in (14) we put $T = T_\nu$, where T_ν was defined just before (12), then $m = Y_0(T_\nu) = \nu$. Let us require in addition that the total number of vertices in the branches be $n - 1$, i.e. $n_1 + n_2 + \cdots n_\nu = n - 1$, then

$$\mathcal{F}_n = \sum_{\nu=1}^{\infty}\sum_{\Sigma n_i = n-1} \mathfrak{M}_{T_\nu}^{n_1\cdots n_\nu}, \qquad n = 2, 3, \cdots, \tag{20}$$

where

$$\mathfrak{M}_{T_i}^{n_1\cdots n_i}\cdot\mathfrak{M}_{T_j}^{m_1\cdots m_j} = 0,$$

unless $i = j$ and $n_1 = m_1$, $n_2 = m_2$, $\cdots$ $n_j = m_j$. By applying P to (20) and using (15) we get the coefficient of Z^n in (19) for $n \geq 2$. This together with the obvious fact that $P_1 = p_0$ completes the proof.

It is worthwhile noticing that by means of the formula of Burman and Lagrange [11] we can solve the recursion formula for P_n in terms of p_0, p_1, $\cdots$, namely

$$P_n = \frac{1}{n!}\left[\frac{d^{(n-1)}}{dw^{n-1}}(f(w))^n\right]_{w=0} = \sum_{\substack{\Sigma\nu_j = n \\ \Sigma j\nu_j = n-1}} \frac{(n-1)!}{\nu_0!\nu_1!\cdots}\, p_0^{\nu_0}p_1^{\nu_1}\cdots. \tag{21}$$

Now if t has n vertices we know from Euler's characteristic that $\sum jY_j(t) = n - 1$. Since $P(t) = \prod p_j^{Y_j(t)}$ we see from (21) that

$$\frac{(n-1)!}{\nu_0!\nu_1!\cdots}, \qquad \Sigma\nu_j = n, \qquad \Sigma j\nu_j = n - 1,$$

is the number of trees in $\mathcal{F}_n$ for which $Y_0(t) = \nu_0$, $Y_1(t) = \nu_1$, $\cdots$.

Evidently $w = \mathcal{P}(z)$ remains a solution of (18) for all z such that $|z| < \alpha$, $|w| < \rho$. In case $p_0 = 0$ the constant 0 solves (18). Hence $\mathcal{P}(z) = 0$ for all z and so $\mathcal{P}(1) = p = 0$. Conversely, if $p = 0$ then $P_1 = p_0 = 0$ which gives

COROLLARY 1. $p = 0$ *if and only if* $p_0 = 0$.

Since we wish to investigate the distribution in $\mathcal{F}$ we shall henceforth assume $p_0 \neq 0$.

Any non-constant function $g(z)$ which has a power series development possessing non-negative coefficients $g(z) = \sum a_\nu z^\nu$, $a_\nu \geq 0$ with a positive radius of convergence R has two properties that are important for us:

(22) $g(z)$ **has a singularity at** R.

(23) If $\sum a_\nu R^\nu$ converges then $\sum a_\nu z_0^\nu$ converges absolutely and uniformly for $|z_0| = R$, and so the series defines a continuous function $g(z)$ there. We have $\lim_{z \to z_0} g(z) = \sum a_\nu z_0^\nu$ as long as the path of approach to z_0 lies in $|z| \leq R$. On the other hand, if as z approaches R through real values below R, $z \to R-$, the limit of $g(z)$ exists then $\sum a_\nu R^\nu$ converges. So if we put $g(R) = \lim_{z \to R-} g(z) = \sum a_\nu R^\nu$ then the meaning is unique even allowing ∞ as a value.

Returning to $\mathcal{P}(z)$, if for $|z| \leq \alpha$ we have $|w| \leq \rho$ where $w = \mathcal{P}(z)$, then

$$z = \frac{w}{f(w)} = \mathcal{P}^{-1}(w), \tag{24}$$

which shows the mapping is schlicht in such a domain and that the image domain cannot contain zeros of $f(w)$. Because of (23) and the fact that $\mathcal{P}(1)$ is finite even if $\alpha = 1$ we see that the mapping is certainly one to one for $|z| \leq 1$.

COROLLARY 2. *p is the smallest root of $f(w) = w$ in $0 \leq w \leq 1$.*

PROOF. (13) shows p is a root in the interval. If for $0 \leq w_0 \leq p$ we have $f(w_0) = w_0$ then by (24) $\mathcal{P}^{-1}(w_0) = 1$.

The following corollary is the well known criterion for extinction

COROLLARY 3. *$p = 1$ if and only if $f'(1) \leq 1$.*

PROOF $p = 1$, $p_0 > 0$, and the convexity of $f(w)$ in $0 \leq w < 1$ guarantee that $(f(w) - 1)/(w - 1)$ is bounded by 1 and is monotonic increasing with w. Hence $f'(1)$ exists and is ≤ 1.

Conversely, if $f'(1) \leq 1$ then either $f'(w)$ is constant ($= p_1 < 1$) in $0 \leq w < 1$ or else it is strictly increasing with w and in either case $f'(w) < 1$. The mean value theorem gives $f(w) > w$ in $0 \leq w < 1$, hence $p = 1$.

Putting $a = \mathcal{P}(\alpha)$ we have the following lemma:

LEMMA 4. $a \leq \rho$

PROOF. We already know that $\mathcal{P}(z)$ has a unique analytic inverse given by (24) for $|\mathcal{P}(z)| < \rho$, but on the other hand $\mathcal{P}'(z) \neq 0$ for $0 \leq z < \alpha$ so this inverse is analytic for $0 \leq w < a$. If we had $a > \rho$ we could continue $f(w)$ analytically by means of (24) along the real axis past its singularity at ρ, but this is impossible.

COROLLARY. *$p = 1$ if and only if $a \geq 1$.*

PROOF. The necessity follows from the monotone behavior of $\mathcal{P}(z)$ for $0 \leq z < \alpha$. Conversely, if $a \geq 1$ then $z = \mathcal{P}^{-1}(1) = 1$.

THEOREM 3. *If $p_0 + p_1 \neq 1$, then*

(25) *α and a are finite;*

(26) $f(a) = a/\alpha$;

(27) *$f'(a) \leq 1/\alpha$ where the strict inequality can hold only if $\alpha = \rho$.*

PROOF. Let $r \geq 2$ be such that $p_r \neq 0$, then for $0 < z < \alpha$, we get from the

functional equation

$$zp_r(\mathcal{P}(z))^r - \mathcal{P}(z) < 0;$$

$$0 < \mathcal{P}(z) < \left(\frac{1}{zp_r}\right)^{1/(r-1)}.$$

By letting $z \to \alpha-$ we see α is finite and $\mathcal{P}(z)$ is bounded. Since $\mathcal{P}(z)$ is monotonic in this region we get $a < \infty$. By letting $z \to \alpha$ in $\mathcal{G}(z, \mathcal{P}(z))$ we get (26). For $0 \leq z \leq \alpha$, $\mathcal{G}_w(z, \mathcal{P}(z)) = zf'(\mathcal{P}(z)) - 1$ is continuous and monotonic increasing with z and is < 0 for z near 0. From the general theorem on implicit functions we know $\mathcal{G}_w(z, \mathcal{P}(z)) \neq 0$ for $|z| < \alpha$, so if we let $z \to \alpha$ we obtain (27).

If $a = \rho$ (27) merely guarantees the finiteness of $f'(\rho)$ and gives an upper bound. One can easily construct an example where $1/\alpha$ is the least upper bound and one where it is not.

But if $a < \rho$ then since $\mathcal{G}(z, w)$ is analytic at (α, a) and $\mathcal{G}(\alpha, a) = 0$ we obtain from the implicit function theorem the strict equality in (27).

COROLLARY. *If* $\alpha = 1$ *then* $a = p = 1$.

PROOF. By (26)

$$f(a) = a = \mathcal{F}(1) = p \leq 1. \tag{28}$$

If $a < \rho$ then $f'(p) = 1$ so $p = 1$ from the convexity of $f(w)$. If $a = \rho$ then $a \geq 1$ which when combined with (28) gives $a = 1$.

The case where $p_0 + p_1 = 1$ escapes Theorem 3 but it is easily examined separately, namely

$$f(w) = p_0 + p_1 w, \qquad p_0 \neq 0,$$

$$\mathcal{P}(z) = \sum_{n=1}^{\infty} p_0 p_1^{n-1} z^n = \frac{p_0 z}{1 - p_1 z}.$$

Hence $p = 1$, $\alpha = 1/p_1$ and $a = \rho = \infty$.

For the practical applications of the theory it is valuable to know some conditions which guarantee $a < \rho$, and thus strict equality in (27). From the foregoing analysis it is evident that one such condition is $\rho = \infty$, i.e. $f(w)$ is an entire function, and another is $f'(1) > 1$. If one has enough information about $f(w)$ to plot its graph for real positive w then the line through the origin tangent to $f(w)$ in the first quadrant touches the curve at the point $(a, a/\alpha)$ from which we determine both a and α.

6. Asymptotic properties of the distributions. If we examine the terms of the sequence $p_0, p_1, \cdots$ we may find that the indices of the non-zero terms are all multiples of some common integer larger than 1. In this case we should expect to have $P_n = 0$ with the same sort of regularity. So let us define q to be the largest integer such that $p_\nu \neq 0$ implies ν is a multiple of q. Clearly we have $q \geq 1$ and $q = 1$ means there is no integer other than 1 which divides the indices of all the non-zero p_ν. Of course, $p_1 \neq 0$ implies $q = 1$ The following theorem establishes an asymptotic estimate for P_n valid for large n, provided

$n - 1$ is a multiple of q, and incidentally shows that $P_n = 0$, if $n - 1$ is not a multiple of q.

THEOREM 4. *If* $a < \rho$ *then*

$$P_n = \begin{cases} q\left(\frac{a}{2\pi\alpha f''(a)}\right)^{\frac{1}{2}} \alpha^{-n} n^{-\frac{1}{2}} + O(\alpha^{-n} n^{-\frac{1}{2}}), & n \equiv 1 \pmod{q}; \\ 0, & n \not\equiv 1 \pmod{q}, \end{cases} \tag{29}$$

i.e. for large $n \equiv 1 \pmod{q}$

$$P_n \sim q\left(\frac{a}{2\pi\alpha f''(a)}\right)^{\frac{1}{2}} \alpha^{-n} n^{-\frac{1}{2}}.$$

PROOF. Let us put $\theta = 2\pi/q$, then for $|w| \leq a$,

$$|f(w)| = \left|\sum_{k=0}^{\infty} p_{kq} w^{kq}\right| \leq \sum_{k=0}^{\infty} p_{kq} |w|^{kq} = f(|w|),$$

and the equality evidently holds if and only if $\arg w$ is an integral multiple of θ. Furthermore, if w is such that $|f(w)| = f(|w|)$ and we put $z = \mathcal{P}^{-1}(w)$ then $w = \mathcal{P}(w/f(w))$ so we get

$$|\mathcal{P}(z)| = \mathcal{P}\left(\frac{|w|}{f(|w|)}\right) = \mathcal{P}\left(\frac{|w|}{|f(w)|}\right) = \mathcal{P}(|z|),$$

hence $P_n = 0$, if $n \not\equiv 1 \pmod{q}$.

For $|z| = \alpha$ and $w = \mathcal{P}(z)$ the point (z, w) satisfies (18) by (23). If we put

$$z_\nu = \alpha e^{i\nu\theta},$$
$$w_\nu = a e^{i\nu\theta}, \qquad \nu = 0, 1, \cdots q - 1,$$

then $w_\nu = \mathcal{P}(z_\nu)$ and

$$\mathcal{G}_w(z_\nu, w_\nu) = z_\nu f'(w_\nu) - 1 = \alpha f'(a) - 1 = 0,$$

so that $z_0, z_1, \cdots z_{q-1}$ are certainly singularities of $\mathcal{P}(z)$. But $f(w)$ is analytic at w_ν and $f(w_\nu) = a/\alpha \neq 0$, so the solution of (18) for z,

$$z = \mathcal{P}^{-1}(w) = \frac{w}{f(w)},$$

is analytic at w_ν. Furthermore

$$\frac{d}{dw}\mathcal{P}^{-1}(w_\nu) = \frac{1 - z_\nu f'(w_\nu)}{f(w_\nu)} = 0,$$

$$\frac{d^2}{dw^2}\mathcal{P}^{-1}(w_\nu) = -\frac{z_\nu f''(w_\nu)}{f(w_\nu)} = -\frac{\alpha^2 f''(a)}{w_\nu} \neq 0,$$

which shows that $\mathcal{P}(z)$ has a branch point of order 1 at each z_ν, i.e. $\mathcal{P}(z)$ is an analytic function of $(z - z_\nu)^{1/2}$ in the neighborhood of (z_ν, w_ν), $\nu = 0, 1, \cdots q - 1$.

For $|z| = \alpha$, $w = \mathcal{P}(z)$ but $z \neq z_\nu$ we obtain

$$|\mathcal{G}_w(z, w)| > 1 - \alpha |f'(w)| > 1 - \alpha f'(|w|) > 1 - \alpha f'(a) = 0,$$

hence $\mathcal{P}(z)$ is an analytic function of z in a certain neighborhood of such a pair (z, w).

By analytic continuation we find a circle of radius $\beta > \alpha$ such that $\mathcal{P}(z)$ is an analytic function of $(z - z_\nu)^{1/2}$ for $|z| \leq \beta$. If we make radial cuts in this circle running outward from each z_ν then in the resulting domain D each of the functions $(z - z_\nu)^{1/2}$ is an analytic function of z hence so is $\mathcal{P}(z)$.

Let Γ be the path consisting of the boundary of D oriented in the positive sense, let γ be the part of Γ lying in the sector $-\pi/q \leq \arg z \leq \pi/q$, and let γ' be that part of γ leading from β to α along the lower lip of the cut at α, thence along the upper lip back to β. Since $\mathcal{P}(z)$ satisfies the relation $\mathcal{P}(e^{i\nu\theta}z) = e^{i\nu\theta}\mathcal{P}(z)$ for $\nu = 0, 1, \cdots q - 1$, we see from Cauchy's formula that

$$P_n = \frac{1}{2\pi i}\int_\Gamma \frac{\mathcal{P}(z)}{z^{n+1}}\,dz = \frac{A}{2\pi i}\int_\gamma \frac{\mathcal{P}(z)}{z^{n+1}}\,dz,$$

where

$$A = \sum_{\nu=0}^{q-1} e^{-i\nu\theta(n-1)} = 0, \qquad n \not\equiv 1 \pmod q;$$
$$= q, \qquad n \equiv 1 \pmod q.$$

Restricting ourselves to $n \equiv 1 \pmod q$ we put

$$\mathcal{P}(z) = a + b(z - \alpha)^{1/2} + c(z - \alpha) + (z - \alpha)^{3/2}\mathcal{Q}(z),$$

where $\mathcal{Q}(z)$ is analytic in D. Then $P_n = B + C$, where

$$B = \frac{q}{2\pi i}\int_\gamma \frac{a + b(z - \alpha)^{\frac{1}{2}} + c(z - \alpha)}{z^{n+1}}\,dz,$$

$$C = \frac{q}{2\pi i}\int_\gamma \frac{(z - \alpha)^{\frac{1}{2}}\mathcal{Q}(z)}{z^{n+1}}\,dz.$$

We find

$$B = \frac{bq}{2\pi i}\int_\Gamma \frac{(z - \alpha)^{\frac{1}{2}}}{z^{n+1}}\,dz + O(\beta^{-n}) = ibq\sqrt{\alpha}(-1)^n\binom{1/2}{n}\alpha^{-n} + O(\beta^{-n});$$

$$|C| = O\left(\int_{\gamma'} \frac{|z - \alpha|^{\frac{1}{2}}}{|z|^{n+1}}\,|dz|\right) = O\left(\left|\int_{\gamma'} \frac{(z - \alpha)^{\frac{1}{2}}}{z^{n+1}}\,dz\right|\right) = O\left(\alpha^{-n}\left|\binom{3/2}{n}\right|\right).$$

The constant b is determined from the equations

$$w - a = b(z - \alpha)^{1/2} + \cdots;$$

$$z - \alpha = -\frac{\alpha^2 f''(a)}{2a}(w - a)^2 + \cdots.$$

Using the fact that

$$\left|\binom{1/2}{n}\right| = (4\pi n^3)^{-1/2} + O(n^{-5/2}),$$

$$\left|\binom{3/2}{n}\right| = O(n^{-5/2}),$$

we finally obtain (29) as desired.

Thus P_n approaches zero a little faster than exponentially with n regardless of whether $p = 1$ or $p < 1$, except for the special case when $\alpha = 1$. In this case it is interesting that, according to the corollary to lemma 4, $p = 1$.

The case where $q \neq 1$ is of no practical importance since one can always bring q back to 1 by making a very small decrease in one of the non-zero p_i and increasing p_1 by the same amount. This can clearly be done so that none of the important characteristics of $f(w)$ is changed appreciably.

7. The limiting distributions of $W(t)$ and $n^{-1}Y_k(t)$ for $t \in \mathcal{F}_n$. Let us momentarily drop the condition $p_0 \neq 0$. The characteristic function of W is

$$\int_{\mathcal{J}} e^{i\theta W}\, dP = f(e^{i\theta}), \tag{30}$$

so that for the rth moment of W we have

$$E(W^r) = \frac{d^{(r)}}{d(i\theta)^r} f(e^{i\theta})\Big|_{\theta=0}, \qquad r = 0, 1, 2, \cdots. \tag{31}$$

For the first and second moments we obtain

$$E(W) = f'(1),$$
$$E(W^2) = f'(1) + f''(1),$$

which shows that the criterion for extinction (Corollary 3 to Theorem 2) may be stated as follows: the multiplicative process is almost certain to expire if and only if $E(W) \leq 1$. From (30) we see that all the moments of W will be finite as soon as $\rho > 1$; but if $\rho = 1$ no general statement can be made, except in case $a = 1$ also, for indeed $a = 1$ implies $\alpha = 1$ so by (31) and (27) $E(W) = f'(1) \leq 1$.

We now reassume $p_0 \neq 0$. Since the variables Z, Y_0, Y_1, $\cdots$ are restricted to $t \in \mathcal{F}$ it is convenient to see what happens to W in $\mathcal{F}$. If we define $g(w) = p^{-1}f(pw)$ then (13) shows $g(w)$ and $g(e^{i\theta})$ are the generating function and characteristic function respectively for W, given $t \in \mathcal{F}$. Thus we see immediately that the first moment of W, given $\mathcal{F}$, is always ≤ 1, and all its moments are finite if $p < 1$.

In case $0 < p < 1$ we may also introduce $h(w)$ defined by

$$f(w) = pg(w) + (1 - p)h(w), \tag{32}$$

then $h(w)$ is obviously the generating function of W, given $\mathcal{G}$. Here the rth moment is finite whenever the rth moment of W is finite. (32) gives

$$P(W = k/\mathcal{G}) = p_k \frac{1 - p^k}{1 - p}, \qquad k = 1, 2, \cdots.$$

It would be interesting to be able to compare this with the corresponding thing for large finite trees and in this connection we have the following theorem:

THEOREM 5. *If $a < \rho$ and $q = 1$,*

$$\lim_{n\to\infty} P(W = k/\mathcal{F}_n) = \alpha k p_k a^{k-1}, \qquad k = 1, 2, \cdots.$$

PROOF. By expanding $zf(e^{i\theta}\mathcal{P}(z))$ in powers of z we obtain

$$zf(e^{i\theta}\mathcal{P}(z)) = \sum_{n=1}^{\infty} \phi_n(\theta) z^n, \tag{33}$$

where

$$\phi_n(\theta) = \int_{\mathcal{F}_n} e^{i\theta W} dP = \sum_{r=1}^{\infty} \sum_{\Sigma n_j = n-1} e^{ir\theta} p_r P_{n_1} P_{n_2} \cdots P_{n_r},$$

so that if $P_n \neq 0$ then $P_n^{-1}\phi_n(\theta)$ is the characteristic function of W, given $\mathcal{F}_n$. From (33) we get

$$\phi_n(\theta) = \frac{1}{2\pi i} \int_\Gamma \frac{f(e^{i\theta}\mathcal{P}(z))}{z^n} dz.$$

Since $a < \rho$ we may expand $f(e^{i\theta}\mathcal{P}(z))$ about the point $\mathcal{P}(z) = a$ and integrate as in the proof of theorem 4, thus

$$P_n^{-1}\phi_n(\theta) = \frac{P_{n-1}}{P_n} e^{i\theta} f'(ae^{i\theta}) + \epsilon_n(\theta).$$

Since $\epsilon_n(\theta) \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} P_n^{-1}\phi_n(\theta) = \alpha e^{i\theta} f'(ae^{i\theta}),$$

the limit function obviously being the characteristic function for the distribution whose generating function is $\alpha w f'(aw)$, from which the theorem follows directly.

Now $\mathcal{P}(z)/p$ is the generating function for Z, given $\mathcal{F}$, and the function solves

$$zg(w) - w = 0 \tag{34}$$

for $|z| < \alpha$. We find for the rth moment of Z

$$E(Z^r/\mathcal{F}) = \frac{d^{(r)}}{d(i\theta)^r} \frac{\mathcal{P}(e^{i\theta})}{p}\bigg|_{\theta=0}, \qquad r = 0, 1, \cdots,$$

hence all the moments are finite as soon as $\alpha > 1$. Since by (34)

$$\frac{dw}{dz} = \frac{g(w)}{1 - zg'(w)} = \frac{w}{z(1 - zg'(w))}, \qquad w = \frac{\mathcal{P}(z)}{p},$$

we obtain for the first moment

$$E(Z/\mathcal{F}) = \frac{\mathcal{P}'(1)}{p} = \frac{1}{1 - g'(1)} = \frac{1}{1 - f'(p)}.$$

In a similar way one can express any moment of Z, provided it is finite, in terms of $f'(p), f''(p)$, etc. If $\alpha = 1$ we see from the corollary to theorem 3 that even the first moment of Z is infinite, except for the special case where $\rho = 1$ and $f'(1) < 1$.

The characteristic function of Y_k, given $\mathcal{F}$, is

$$\psi_k(\theta) = 1/p \int_{\mathcal{F}} e^{i\theta Y_k} dP = 1/p \sum_{n=1}^{\infty} \psi_{kn}(\theta),$$

where by (21)

$$\psi_{kn}(\theta) = \int_{\mathfrak{F}_n} e^{i\theta\nu_k}\, dP = \sum_{\substack{\Sigma\nu_j = n \\ \Sigma j\nu_j = n-1}} \frac{(n-1)!}{\nu_0!\,\nu_1!\cdots}\, e^{i\nu_k\theta}\, p_0^{\nu_0} p_1^{\nu_1} \cdots .$$

Thus, if $P_n \not\equiv 0$, $P_n^{-1}\psi_{kn}(\theta)$ is the characteristic function of Y_k, given $\mathfrak{F}_n$. If $p_k = 0$ then $\psi_k(\theta) = 1$. If $p_k \not\equiv 0$ put $p_k = e^{q_k}$ then

$$\frac{\partial^{(r)}}{\partial q_k^r}\, \mathcal{P}(q) = \sum_{n=1}^{\infty} \psi_{kn}^{(r)}(0) z^n, \tag{35}$$

hence

$$\frac{1}{p}\frac{\partial^{(r)}}{\partial q_k^r}\, \mathcal{P}(1) = E(Y_k^r/\mathfrak{F}),$$

which shows that all moments of Y_k are finite if $\alpha > 1$. Let us put $w = \mathcal{T}(z)$, for short, then, by (18),

$$\frac{\partial w}{\partial q_k} = \frac{z p_k w^k}{1 - z f'(w)} = z^2 p_k w^{k-1} \mathcal{P}'(z), \tag{36}$$

which gives for the first moment of Y_k,

$$E(Y_k/\mathcal{F}) = \frac{p_k p^{k-1}}{1 - f'(p)} = p_k p^{k-1} E(Z/\mathfrak{F}),$$

which is to be expected since $p_k p^{k-1}$ plays the same role in $\mathfrak{F}$ that p_k plays in $\mathcal{T}$. We may also expect that for $\iota \in \mathfrak{F}_n$, $n^{-1}Y_k$ should be closely related to p_k. This question is settled by the following theorem:

THEOREM 6. *If $a < \rho$ and $q = 1$ then for x real*

$$\lim_{n\to\infty} P(n^{-1}Y_k < x/\mathfrak{F}_n) = \begin{cases} 1, & \text{if } x \geq \alpha p_k a^{k-1}; \\ 0, & \text{if } x < \alpha p_k a^{k-1}. \end{cases}$$

PROOF. We intend to estimate the rth moment of $n^{-1}Y_k$ for $\iota \in \mathfrak{F}_n$ and n very large from (35) by means of the contour integral

$$E(n^{-r} Y_k^r/\mathfrak{F}_n) = \frac{1}{2\pi i n^r P_n} \int_{\mathfrak{r}} \frac{\partial^{(r)}}{\partial q_k^r}\, \mathcal{P}(z)\, \frac{dz}{z^{n+1}}. \tag{37}$$

So let us put

$$w = \mathcal{P}(z), \qquad w_r^{(s)} = \frac{\partial^{(r+s)}}{\partial q_k^r \partial z^s}\, w, \qquad r, s = 0, 1, \cdots,$$

then by (36) $w_1 = z^2 p_k w^{k-1} w^{(1)}$ and by Leibnitz formula, provided $k \not\equiv 0$,

$$w_r z^2 p_k \sum_{\substack{\Sigma\nu_i = r-1 \\ \nu_i \geq 0}} \frac{(r-1)!}{\nu_0!\,\nu_1! \cdots \nu_k!}\, w_{\nu_1} w_{\nu_2} \cdots w_{\nu_{k-1}} w_{\nu_k}^{(1)}. \tag{38}$$

The principal contribution to the integral in (37) will come from the term of (38) which has the largest size for z near α. If we put $\zeta = (z - \alpha)^{1/2}$ then w is regular at $\zeta = 0$ and so is the constant p_k. Let's assume that w_ν has a pole of order $2\nu - 1$ at $\zeta = 0$ for $\nu = 1, 2, \cdots r - 1$, which is clearly true for $\nu = 1$. Then if s is the number of $\nu_1, \nu_2, \cdots \nu_{k-1}$ which are $= 0$, the order of the pole of the general term of (38) at $\zeta = 0$ is

$$\sum_{i=1}^{k-1} (2\nu_i - 1) + s + 2\nu_k + 1 = 2(r - \nu_0) - (k - s),$$

which has the maximum value $2r - 1$ if and only if $\nu_0 = \nu_1 = \cdots \nu_{k-1} = 0$, $\nu_k = r - 1$. Hence

$$w_r = z^2 p_k w^{k-1} w_{r-1}^{(1)} + \zeta^{2-2r}\mathcal{R}_1(\zeta), \tag{39}$$

where $\mathcal{R}(\zeta)$ is a regular function of ζ at zero. For $k = 0$ the formula (38) is not correct but it is easy to see directly that (39) is correct for $k = 0$. If we derive (39) with respect to z and put $r - 1$ for r we obtain

$$w_{r-1}^{(1)} = z^2 p_k w^{k-1} w_{r-2}^{(2)} + \zeta^{2-2r}\mathcal{R}_2(\zeta),$$

hence

$$w_r = (z^2 p_k w^{k-1})^r w^{(r)} + \zeta^{2-2r}\mathcal{R}_3(\zeta).$$

Substituting in (37) and estimating in a manner similar to that employed previously we obtain

$$E(n^{-r} Y_k^r/\mathcal{F}_n) = \frac{(p_k a^{k-1})^r}{2\pi i n^r p_n} \int_\Gamma \frac{\mathcal{P}^{(r)}(z)}{z^{n-2r+1}}\, dz + \int_\Gamma \frac{(z - \alpha)^{1-r}\mathcal{R}_4((z - \alpha)^{1/2})}{2\pi i n^r p_n z^{n+1}}\, dz$$

$$= (p_k a^{k-1})^r \frac{P_{n-r}}{P_n} \frac{(n - r)(n - r - 1) \cdots (n - 2r + 1)}{n^r} + O(n^{-1/2}),$$

and finally

$$\lim_{n\to\infty} E(n^{-r} Y_k^r/\mathcal{F}_n) = (\alpha p_k a^{k-1})^r.$$

The limit of the rth moment is itself the rth moment of the distribution on the real line which has all its mass at the point $\alpha p_k a^{k-1}$. Since this distribution is uniquely determined by its moments, a well known theorem [7] enables us to conclude that our sequence of distributions has this distribution as limit and this is equivalent to what is claimed by the theorem.

It is important to notice that if we put the mass $\alpha p_k a^{k-1}$ at the point k this determines a distribution on the real line because of (26).

8. The estimation of p. If we wish to estimate p when we know $p \neq 0$, we may obtain an estimate from the knowledge of $f(w)$ in $0 < w < 1$, using the method of iteration. That is we choose a function $G(w)$ such that $G(p) = p$ and $|G(w) - p| < |w - p|$ for $0 < w < 1$. Then if for any w_0 in the open

interval we compute successively $w_1, w_2, \cdots$, where $w_{n+1} = G(w_n)$ for $n \geq 0$, we are sure that w_n converges exponentially to p as $n \to \infty$.

Obviously $f(w)$ itself has the properties of $G(w)$ but we achieve faster convergence towards p using Newton's method, that is if we put

$$f_1(w) = f(w) - w,$$
$$G(w) = w - \frac{f_1(w)}{f_1'(w)}. \tag{40}$$

If for some reason we expect p to be close to 1 then it is better to put

$$f_2(w) = \frac{f(w) - w}{w - 1},$$

and use $f_2(w)$ in (40) instead of $f_1(w)$, for then we may choose $w_0 = 1$.

Let us put $f'(1) = 1 + \epsilon$, $\epsilon > 0$ then

$$f_2(1 + h) = \frac{f(1 + h) - 1}{h} - 1 \to \epsilon, \qquad h \to 0;$$

$$f_2'(1 + h) = \lim_{k \to 0} \left(\frac{f(1 + h + k) - 1}{k(h + k)} - \frac{f(1 + h) - 1}{kh} \right),$$

$$f_2'(1) = \lim_{h \to 0} \left(\frac{f(1 + 2h) - 2f(1 + h) + 1}{2h^2} \right) = \frac{f''(1)}{2}.$$

Hence

$$p \approx w_1 = 1 - \frac{2\epsilon}{f''(1)}. \tag{41}$$

This result was previously established by Kolmogoroff [7].

The following two simple examples display the results of the general theory.

EXAMPLE 1. We take $f(w) = p_0 + p_1 w + p_2 w^2$ where $p_0 + p_1 + p_2 = 1$ and $p_0, p_2 > 0$. We have $\rho = \infty$. From the equations (26) and (27),

$$f(a) = p_0 + p_1 a + p_2 a^2 = \frac{a}{\alpha},$$

$$f'(a) = p_1 + 2p_2 a = \frac{1}{\alpha},$$

we obtain easily

$$a = \sqrt{p_0 p_2^{-1}}, \qquad \alpha^{-1} = p_1 + 2\sqrt{p_0 p_2},$$

and it is evident that $a \geq 1$ is equivalent to $p_0 \geq p_2$ is equivalent to $f'(1) = p_1 + 2p_2 \leq 1$. Now

$$\mathcal{G}(z, w) = zp_0 + (zp_1 - 1)w + zp_2 w^2,$$

hence

$$\mathcal{P}(z) = \frac{1 - zp_1 - \sqrt{(1 - zp_1)^2 - 4z^2p_0p_2}}{2zp_2}, \tag{42}$$

the choice of the sign of the radical being determined by letting $z \to 0$.

$$p = \frac{p_0 + p_2 - \sqrt{(p_0 - p_2)^2}}{2p^2} = \begin{cases} 1, & p_0 \geq p_2; \\ p_0 p_2^{-1}, & p_0 < p_2. \end{cases}$$

In the case $p_1 > 0$ we have $q = 1$ and then by (21)

$$p_n = \sum_{\substack{\nu_0+\nu_1+\nu_2=n \\ \nu_1+2\nu_2=n-1}} \frac{(n-1)!}{\nu_0!\,\nu_1!\,\nu_2!}\, p_0^{\nu_0} p_1^{\nu_1} p_2^{\nu_2},$$

which can also be obtained by expansion of (42) according to powers of z. From (29) we get

$$p_n \sim \sqrt{\frac{1}{4\pi}\,(p_1\sqrt{p_0p_2^{-3}} + 2p_0p_2^{-1})}\,(p_1 + 2\sqrt{p_0p_2})^n\, n^{-3/2}.$$

In the case $p_1 = 0$ we have $q = 2$ and obtain from (42) or from (29)

$$\mathcal{P}(z) = \sum_{\nu=1}^{\infty} (-1)^{\nu+1} \binom{1/2}{\nu} 2^{2\nu-1} p_0^{\nu} p_2^{\nu-1} z^{2\nu-1}$$

$$= \sum_{\nu=1}^{\infty} \frac{(2\nu-2)!}{\nu!\,(\nu-1)!}\, p_0^{\nu} p_2^{\nu-1} z^{2\nu-1},$$

which shows

$$P_n = \begin{cases} 0, & n = 2\nu; \\ \dfrac{(2\nu-2)!}{\nu!(\nu-1)!}\, p_0^{\nu} p_2^{\nu-1}, & n = 2\nu - 1. \end{cases}$$

By direct use of Stirling's formula or from (29) we get

$$P_{2\nu-1} \sim \frac{1}{p_2}\sqrt{\frac{2}{\pi}}\, 2^{2\nu-1} (p_0p_2)^{\nu} (2\nu-1)^{3/2}.$$

EXAMPLE 2. We take $f(w) = e^{\lambda(w-1)}$, $\lambda > 0$, so that W has a Poisson distribution. Then $\rho = \infty$, $q = 1$, and we get from (26) and (27)

$$f(a) = e^{\lambda(a-1)} = a/\alpha,$$

$$f'(a) = \lambda e^{\lambda(a-1)} = 1/\alpha,$$

$$a = 1/\lambda, \qquad \alpha = e^{\lambda-1}/\lambda.$$

Clearly we have $a \geq 1$ if and only if $\lambda \leq 1$ and in this case 1 is evidently the only solution for w of $e^{\lambda(w-1)} = w$, hence $p = 1$. On the other hand if $\lambda < 1$

then (41) gives $p = 1 - 2(\lambda - 1)\lambda^{-2}$. By (21) we get

$$P_n = \frac{(n\lambda)^{n-1}}{n!} e^{-n\lambda},$$

and by direct use of Stirling's formula or from (29) we get

$$P_n \sim \sqrt{\frac{1}{2\pi}}\, e^{n(1-\lambda)} \lambda^{n-1} n^{-3/2}.$$

REFERENCES

[1] R. A. Fisher, *The Genetical Theory of Natural Selection*, Oxford, The Clarendon Press 1930.

[2] A. J. Lotka, *Theorie Analytique des Associations Biologiques* 2, Hermann and Co., Paris, 1939.

[3] D. Hawkins and S. Ulam, *Theory of Multiplicative Processes* I, MDDC-287, 1944.

[4] T. E. Harris, "Some theorems on the Bernoullian multiplicative process," *thesis, doctor of philosophy*, Princeton University, 1947.

[5] A. M. Yaglom, "Certain limit theorems of the theory of branching random processes," *Doklady Akad. Nauk. SSSR(N. S.)* Vol. 56 (1947), 795–798.

[6] D. G. Kendall, "On the generalized "birth-and-death" process," *Annals of Math. Stat.*, Vol. 19 (1948).

[7] A. Kolmogoroff, "Zur Lösung einer biologischen Aufgabe," *Mitt. Forsch.-Inst. Math. u. Mech. Univ. Tomsk*, Vol. 2 (1938), pp. 1–6.

[8] L. Pontrjagin, *Topological Groups*, Princeton Univ. Press, 1946.

[9] J. von Neumann, *Functional Operators*, mimeographed notes, Institute for Advanced Study, 1933–35.

[10] A. Kolmogoroff, *Grundbergriffe der Wahrscheinlichkeitsrechnung*, Chelsea Publishing Co., New York, 1946.

[11] A. Hurwitz and R. Courant, *Funktionentheorie*, Springer, Berlin, 1929.

[12] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. I, Griffin Co., London, 1943.

APPLICATION OF THE RADON-NIKODYM THEOREM TO THE THEORY OF SUFFICIENT STATISTICS[1]

By Paul R. Halmos[2] and L. J. Savage

University of Chicago

Summary. The body of this paper is written in terms of very general and abstract ideas which have been popular in pure mathematical work on the theory of probability for the last two or three decades. It seems to us that these ideas, so fruitful in pure mathematics, have something to contribute to mathematical statistics also, and this paper is an attempt to illustrate the sort of contribution we have in mind. The purpose of generality here is not to solve immediate practical problems, but rather to capture the logical essence of an important concept (sufficient statistic), and in particular to disentangle that concept from such ideas as Euclidean space, dimensionality, partial differentiation, and the distinction between continuous and discrete distributions, which seem to us extraneous.

In accordance with these principles the center of the stage is occupied by a completely abstract sample space—that is a set X of objects x, to be thought of as possible outcomes of an experimental program, distributed according to an unknown one of a certain set of probability measures. Perhaps the most familiar concrete example in statistics is the one in which X is n dimensional Cartesian space, the points of which represent n independent observations of a normally distributed random variable with unknown parameters, and in which the probability measures considered are those induced by the various common normal distributions of the individual observations.

A statistic is defined, as usual, to be a function T of the outcome, whose values, however, are not necessarily real numbers but may themselves be abstract entities. Thus, in the concrete example, the entire set of n observations, or, less trivially, the sequence of all sample moments about the origin are statistics with values in an n dimensional and in an infinite dimensional space respectively. Another illuminating and very general example of a statistic may be obtained as follows. Suppose that the outcomes of two not necessarily statistically independent programs are thought of as one united outcome—then the outcome T of the first program alone is a statistic relative to the united program. A technical measure theoretic result, known as the Radon-Nikodym theorem, is important in the study of statistics such as T. It is, for example, essential to the very definition of the basic concept of conditional probability of a subset E of X given a value y of T.

The statistic T is called sufficient for the given set $\mathfrak{M}$ of probability measures

[1] This paper was the basis of a lecture delivered upon invitation of the Institute at the meeting in Chicago on December 30, 1947.

[2] Fellow of the John Simon Guggenheim Memorial Foundation.

if (somewhat loosely speaking) the conditional probability of a subset E of X given a value y of T is the same for every probability measure in $\mathfrak{M}$. It is, for instance, well known that the sample mean and variance together form a sufficient statistic for the measures described in the concrete example.

The theory of sufficiency is in an especially satisfactory state for the case in which the set $\mathfrak{M}$ of probability measures satisfies a certain condition described by the technical term *dominated*. A set $\mathfrak{M}$ of probability measures is called dominated if each measure in the set may be expressed as the indefinite integral of a density function with respect to a fixed measure which is not itself necessarily in the set It is easy to verify that both classical extremes, commonly referred to as the discrete and continuous cases, are dominated.

One possible formulation of the principal result concerning sufficiency for dominated sets is a direct generalization to the abstract case of the well known Fisher-Neyman result: T is sufficient if and only if the densities can be written as products of two factors, the first of which depends on the outcome through T only and the second of which is independent of the unknown measure. Another way of phrasing this result is to say that T is sufficient if and only if the likelihood ratio of every pair of measures in $\mathfrak{M}$ depends on the outcome through T only. The latter formulation makes sense even in the not necessarily dominated case but unfortunately it is not true in that case. The situation can be patched up somewhat by introducing a weaker notion called pairwise sufficiency.

In ordinary statistical parlance one often speaks of a statistic sufficient for some of several parameters. The abstract results mentioned above can undoubtedly be extended to treat this concept.

1. Basic definitions and notations. A measurable space $(X, \mathcal{S})$ is a set X and a σ-algebra $\mathcal{S}$ of subsets of X.[3] If $(X, \mathcal{S})$ and $(Y, \mathcal{T})$ are measurable spaces and if T is a transformation from X into Y (or, in other words, if T is a function with domain X and range in Y), then T is *measurable* if, for every F in $\mathcal{T}$, $T^{-1}(F) \in \mathcal{S}$. If Y is a Borel set in a finite dimensional Euclidean space, then we shall always understand that $\mathcal{T}$ is the class of all Borel subsets of Y, and the measurability of a function f from X to Y will be expressed by the notation $f(\epsilon)\ \mathcal{S}$.

Throughout most of what follows it will be assumed that $(X, \mathcal{S})$ and $(Y, \mathcal{T})$ are fixed measurable spaces and that T is a measurable transformation (also called a *statistic*) from X *onto* Y. A helpful example to keep in mind is the Cartesian plane in the role of X, its horizontal coordinate axis in the role of Y, and perpendicular projection from X onto Y in the role of T.

The following notations will be used. If g is a point function on Y (with arbitrary range), then gT is the point function on X defined by $gT(x) = g(T(x))$. If μ is a set function (with arbitrary range) on $\mathcal{S}$, then μT^{-1} is the set function

[3] A σ-algebra is a non empty class $\mathcal{S}$ of sets, closed under the formation of complements and countable unions. If $(X, \mathcal{S})$ is a measurable space, the sets of $\mathcal{S}$ will be called the measurable sets of X

on T defined by $\mu T^{-1}(F) = \mu(T^{-1}(F))$. The class of all sets of the form $T^{-1}(F)$, with $F \in \mathcal{T}$, will be denoted by $T^{-1}(\mathcal{T})$. the characteristic function of a set A (in any space) will be denoted by χ_A.

LEMMA 1. *If g is any function on Y and A is any set in the range of g, then*

$$\{x: gT(x) \in A\} = T^{-1}(\{y: g(y) \in A\});$$

hence, in particular, $\chi_{T^{-1}(F)} = \chi_F T$ for every subset F of Y.[4]

PROOF. The following statements are mutually equivalent: (a) $x_0 \in \{x: gT(x) \in A\}$, (b) $g(T(x_0)) \in A$, (c) if $y_0 = T(x_0)$, then $g(y_0) \in A$, and (d) $T(x_0) \in \{y: g(y) \in A\}$. The equivalence of the first and last ones of these statements is exactly the assertion of the lemma.

We shall have frequent occasion to deal with functions on X which are induced by measurable functions on Y; the following result is a useful and direct structural characterization of such functions.

LEMMA 2. *If f is a real valued function on X, then a necessary and sufficient condition that there exist a measurable function g on Y such that $f = gT$ is that $f \in T^{-1}(\mathcal{T})$; if such a function g exists, then it is unique.*[5]

PROOF. The necessity of the condition is clear. To prove sufficiency, suppose that $f \in T^{-1}(\mathcal{T})$, $y_0 \in Y$, and write $X_0 = T^{-1}(\{y_0\})$. Suppose $x_0 \in X_0$ and write $E = \{x: f(x) = f(x_0)\}$. Since $f \in T^{-1}(\mathcal{T})$, there is a set F in $\mathcal{T}$ such that $E = T^{-1}(F)$. Since $x_0 \in E$, it follows that $y_0 \in F$ and therefore that

$$X_0 = T^{-1}(\{y_0\}) \subset T^{-1}(F) = E.$$

In other words f is a constant on X_0 and consequently the equation $g(y_0) = f(x_0)$ unambiguously defines a function g on Y. The facts that $f = gT$ and that g is measurable are clear; the uniqueness of g follows from the fact that T maps X onto Y.

2. Measures and their derivatives. A *measure* is a real valued, non negative, finite (and therefore bounded), countably additive function on the measurable sets of a measurable space.[6] An integral whose domain of integration is not indicated is always to be extended over the whole space. If the symbol $[\mu]$, pronounced "modulo μ", follows an assertion concerning the points x of X, it is to be understood that the set E of those points for which the assertion is not true is such that $E \in \mathcal{S}$ and $\mu(E) = 0$. Thus, for instance, if f and g are functions (with arbitrary range) on X, then $f = g$ $[\mu]$ means that

[4] The symbol $\{— : —\}$ stands for the set of all those objects named before the colon which satisfy the condition stated after it.

[5] The notation $f \in T^{-1}(\mathcal{T})$ means of course that f is a measurable function not only on the measurable space $(X, \mathcal{S})$ but also on the measurable space $(X, T^{-1}(\mathcal{T}))$. The restriction to real valued functions is inessential and is made only in order to avoid the introduction of more notation.

[6] Although most of the measures occurring in the applications of our theory are *probability measures* (i.e. measures whose value for the whole space is 1), the consideration of probability measures only is, in many of the proofs in the sequel, both unnecessary and insufficient.

$\mu(\{x: f(x) \neq g(x)\}) = 0$. Similarly, if f is a real valued function on X, then $f\ (\epsilon)\ T^{-1}(T)\ [\mu]$ means that there exists a real valued function g on X such that $g\ (\epsilon)\ T^{-1}(T)$ and $f = g\ [\mu]$.

If μ and ν are two measures on S, ν is *absolutely continuous* with respect to μ, in symbols $\nu \ll \mu$, if $\nu(E) = 0$ for every measurable set E for which $\mu(E) = 0$. The measures μ and ν are *equivalent*, in symbols $\mu \equiv \nu$, if simultaneously $\mu \ll \nu$ and $\nu \ll \mu$.[7] One of the most useful results concerning absolute continuity is the Radon-Nikodym theorem, which may be stated as follows.[8]

A necessary and sufficient condition that $\nu \ll \mu$ is that there exist a non negative function f on X such that

$$\nu(E) = \int_E f(x)\ d\mu(x)$$

for every E in S. The function f is unique in the sense that if also

$$\nu(E) = \int_E g(x)\ d\mu(x)$$

for every E in S, then $f = g\ [\mu]$. If $\nu(E) \leqq \mu(E)$ for every E in S, then $0 \leqq f(x) \leqq 1\ [\mu]$.

It is customary and suggestive to write $f = d\nu/d\mu$. Since $d\nu/d\mu$ is determined only to within a set for which μ vanishes, it follows that in a relation of the form

$$\frac{d\nu}{d\mu}\ (\epsilon)\ T^{-1}(T)\ [\mu]$$

the symbol $[\mu]$ is superfluous and may be omitted.

For typographical and heuristic reasons it is convenient sometimes to write the relation $f = d\nu/d\mu$ in the form $d\nu = f d\mu$; all the properties of Radon-Nikodym derivatives which are suggested by the well known differential formalism correspond to true theorems. Some of the ones that we shall make use of are trivial (e.g. $d\nu_1 = f_1 d\mu$ and $d\nu_2 = f_2 d\mu$ imply $d(\nu_1 + \nu_2) = (f_1 + f_2)d\mu$), while others are well known facts in integration theory (e.g. (i) $d\lambda = f d\nu$ and $d\nu = g d\mu$ imply $d\lambda = fg d\mu$, and (ii) $d\nu = f d\mu$ and $d\mu = g d\nu$ imply $fg = 1\ [\mu]$).

We conclude this section with a simple but useful result concerning the transformations of integrals.

LEMMA 3. *If g is a real valued function on Y and μ is a measure on S, then*

$$\int_F g(y)\, d\mu T^{-1}(y) = \int_{T^{-1}(F)} gT(x)\ d\mu(x)$$

for every F in T, in the sense that if either integral exists, then so does the other and the two are equal.

---

[7] It is clear that the relation of equivalence is reflexive, symmetric, and transitive, and hence deserves its name.

[8] For a proof of the Radon-Nikodym theorem and similar facts concerning the measure and integration theory which we employ, see S. Saks, *Theory of the Integral*, Warszawa—Lwów, 1937.

PROOF. Replacing g by $g\chi_F$ we see that it is sufficient to consider the case $F = Y$. The proof for this case follows from the observation that every approximating sum

$$\Sigma_i\, g(y_i)\mu T^{-1}(F_i)$$

of $\int g d\mu T^{-1}$ is also an approximating sum

$$\Sigma_i\, gT(x_i)\mu(E_i)$$

of $\int gT d\mu$, and conversely.[9]

3. Conditional probabilities and expectations. LEMMA 4. *If μ and ν are measures on S such that $\nu \ll \mu$, then $\nu T^{-1} \ll \mu T^{-1}$.*

PROOF. If $F \in T$ and $0 = \mu T^{-1}(F) = \mu(T^{-1}(F))$, then

$$0 = \nu(T^{-1}(F)) = \nu T^{-1}(F).^{10}$$

Lemma 4 is the basis of the definition of a concept of great importance in probability theory. If μ is a measure on S and f is a non negative integrable function on X, then the measure ν defined by $d\nu = f d\mu$ is absolutely continuous with respect to μ. It follows from Lemma 4 that νT^{-1} is absolutely continuous with respect to μT^{-1}; we write $d\nu T^{-1} = g d\mu T^{-1}$. The function value $g(y)$ is known as the *conditional expectation* of f given y (or given that $T(x) = y$), we shall denote it by $e_\mu(f \mid y)$. If $f = \chi_E$ is the characteristic function of a set E in S, then $e_\mu(f \mid y)$ is known as the *conditional probability* of E given y; we shall denote it by $p_\mu(E \mid y)$.[11]

The abstract nature of these definitions makes an intuitive justification of them desirable. Observe that since $\nu T^{-1}(F) = \nu(T^{-1}(F)) = \int_{T^{-1}(F)} f(x)\, d\mu(x)$, the defining equation of $e_\mu(f \mid y)$, written out in full detail, takes the form

$$\int_{T^{-1}(F)} f(x)\, d\mu(x) = \int_F e_\mu(f \mid y)\, d\mu T^{-1}(y), \qquad F \in T.$$

---

[9] It is of interest to observe that either side of the equation in Lemma 3 may be obtained from the other by the formal substitution $y = T(x)$. A special case of this lemma is the celebrated and often misunderstood assertion that the expectation of a random variable is equal to the first moment of its distribution function.

[10] That the converse of Lemma 4 is not true is shown by the following example. Let X be the unit square, let Y be the unit interval, and let T be the perpendicular projection from X onto Y. Let μ be ordinary (Borel-Lebesgue) measure and let ν be linear measure on the intersection of X with, say, the horizontal line whose ordinate is $\frac{1}{2}$. Clearly ν is not absolutely continuous with respect to μ, but $\nu T^{-1} = \mu T^{-1}$.

[11] Definitions in this form were first proposed by A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933. With a slight amount of additional trouble, conditional expectation could be defined for more general functions, but only the non negative case will occur in our applications.

If $f = \chi_E$, then this equation becomes the defining equation of $p_\mu(E \mid y)$:

$$\mu(E \cap T^{-1}(F)) = \int_F p_\mu(E \mid y)\, d\mu T^{-1}(y), \qquad F \in \mathcal{T}.^{12}$$

The customary definition of "the conditional probability of E given that $T(x) \in F$" is $\mu(E \cap T^{-1}(F))/\mu(T^{-1}(F))$, (assuming that the denominator does not vanish). Since $\mu(T^{-1}(F)) = \mu T^{-1}(F)$, we have

$$\frac{\mu(E \cap T^{-1}(F))}{\mu(T^{-1}(F))} = \frac{1}{\mu T^{-1}(F)} \int_F p_\mu(E \mid y)\, d\mu T^{-1}(y).$$

It is now formally plausible that if "F shrinks to a point y," then the left side of the last written equation should tend to the conditional probability of E given y and the right side should tend to the integrand $p_\mu(E \mid y)$. The use of the Radon-Nikodym differentiation theorem is a rigorous substitute for this rather shaky difference quotient approach.

Since $p_\mu(E \mid y)$ is determined, for each E, only to within a set for which μT^{-1} vanishes, it would be too optimistic to expect that, for each y, it behaves, regarded as a function of E, like a measure. It is, however, easy to prove that

(i) $p_\mu(X \mid y) = 1$ $[\mu T^{-1}]$,

(ii) $0 \leqq p_\mu(E \mid y) \leqq 1$ $[\mu T^{-1}]$,

(iii) if $\{E_n\}$ is a disjoint sequence of measurable sets, then $p_\mu(\bigcup_{n=1}^{\infty} E_n \mid y) = \sum_{n=1}^{\infty} p_\mu(E_n \mid y)$ $[\mu T^{-1}]$.[13]

The exceptional sets of measure zero depend in general on E in (ii) and on the particular sequence $\{E_n\}$ in (iii). It is interesting to observe that, despite the fact that μ need not be a probability measure, p_μ turns out always to have the normalization property (i). It is natural to ask whether or not the indeterminacy of $p_\mu(E \mid y)$ may be resolved, for each E, in such a way that the resulting function is a measure for each y, except possibly for a fixed set of y's on which μT^{-1} vanishes. Doob[14] has shown that this is the case when X is the real line; in the general case such a resolution is impossible. Fortunately, however, conditional probabilities are sufficiently tractable for most practical and theoretical purposes, and the requirement that they should behave like probability measures in the strict sense described above is almost never needed.

---

[12] We observe that it is not sufficient to require this for $F = Y$ only, i.e. to require $\mu(E) = \int p_\mu(E \mid y)\, d\mu T^{-1}(y)$. This special equation is satisfied by many functions which do not deserve the name conditional probability; e.g. it is satisfied by $p_\mu(E \mid y) =$ constant $= \mu(E)/\mu T^{-1}(Y)$.

[13] See J. L. Doob, "Stochastic processes with an integral-valued parameter," *Am. Math. Soc. Trans.*, Vol. 44 (1938), pp. 95–98.

[14] See Doob, *loc. cit.* Doob asserts the theorem in much greater generality, but his proof is incorrect. The error in the proof and a counterexample to the general theorem were communicated to us by J. Dieudonné in a letter dated September 4, 1947. Doob's proof is valid for more general spaces than the real line (e.g. for finite dimensional Euclidean spaces and for compact metric spaces). The details of Dieudonné's counterexample will appear in a forthcoming book (entitled *Measure theory*) by Halmos.

We conclude this section with two easy but useful results which might also serve as illustrations of the method of finding conditional probabilities and expectations in certain special cases.

LEMMA 5. *If μ is a measure on S, if g is a non negative function on Y, integrable with respect to μT^{-1}, and if ν is the measure on S defined by $d\nu = gTd\mu$, then $d\nu T^{-1} = gd\mu T^{-1}$, or, equivalently, $e_\mu(gT \mid y) = g(y)$ $[\mu T^{-1}]$.*

PROOF. From $\nu(E) = \int_E gT(x)\, d\mu(x)$ and Lemma 3 it follows that

$$\nu T^{-1}(F) = \nu(T^{-1}(F)) = \int_F g(y)\, d\mu T^{-1}(y).$$

LEMMA 6. *If μ is a measure on S, if f and g are non negative functions on X and Y respectively, and if f, gT, and $f \cdot gT$ are all integrable with respect to μ, then*

$$e_\mu(f \cdot gT \mid y) = e_\mu(f \mid y) \cdot g(y) \quad [\mu T^{-1}].$$

Hence, in particular, if $F \in \mathcal{T}$, then

$$p_\mu(E \cap T^{-1}(F) \mid y) = p_\mu(E \mid y)\chi_F(y) \quad [\mu T^{-1}]$$

for every E in S.

PROOF. If $d\nu = fd\mu$, then, by definition of e_μ, $\nu T^{-1}(F) = \int_F e_\mu(f \mid y)\, d\mu T^{-1}(y)$. Applications of Lemmas 3 and 5 yield

$$\int_F e_\mu(f \mid y)g(y)\, d\mu T^{-1}(y) = \int_F g(y)\, d\nu T^{-1}(y) = \int_{T^{-1}(F)} gT(x)\, d\nu(x)$$

$$= \int_{T^{-1}(F)} f(x)gT(x)\, d\mu(x) = \int_F e_\mu(f \cdot gT \mid y)\, d\mu T^{-1}(y),$$

and therefore the desired conclusion follows from the uniqueness assertion of the Radon-Nikodym theorem.

4. Dominated sets of measures. In many statistical situations it is necessary to consider simultaneously several measures on the same σ-algebra. The concept of absolute continuity is easily extended to sets of measures. If $\mathfrak{M}$ and $\mathfrak{N}$ are two sets of measures on S and if, for every set E in S, the vanishing of $\mu(E)$ for every μ in $\mathfrak{M}$ implies the vanishing of $\nu(E)$ for every ν in $\mathfrak{N}$, then we shall call $\mathfrak{N}$ absolutely continuous with respect to $\mathfrak{M}$ and write $\mathfrak{N} \ll \mathfrak{M}$. If $\mathfrak{N} \ll \mathfrak{M}$ and $\mathfrak{M} \ll \mathfrak{N}$, the sets $\mathfrak{M}$ and $\mathfrak{N}$ are called equivalent and we write $\mathfrak{M} \equiv \mathfrak{N}$. If, in particular, $\mathfrak{M}$ contains exactly one measure μ, $\mathfrak{M} = \{\mu\}$, the abbreviated notations $\mathfrak{N} \ll \mu$, $\mu \ll \mathfrak{N}$, and $\mu \equiv \mathfrak{N}$, will be employed for $\mathfrak{N} \ll \mathfrak{M}$, $\mathfrak{M} \ll \mathfrak{N}$, and $\mathfrak{M} \equiv \mathfrak{N}$, respectively.

A set $\mathfrak{M}$ of measures on S will be called *dominated* if there exists a measure λ on S (not necessarily in $\mathfrak{M}$) such that $\mathfrak{M} \ll \lambda$. In applications there frequently occur sets of measures which are dominated in a sense apparently weaker than the one just defined—weaker in that the measure λ, which may for instance be

Lebesgue measure on the Borel sets of a finite dimensional Euclidean space, is not necessarily finite. It is easy to see, however, that whenever λ has the property (possessed by Lebesgue measure) that the space X is the union of countably many sets of finite measure, then a finite measure equivalent to λ exists and the two possible definitions of domination coincide.

The following result on dominated sets of measures may be found to have some interest of its own and will be applied in the sequel.

LEMMA 7. *Every dominated set of measures has an equivalent countable subset.*

PROOF. Let $\mathfrak{M}$ be a dominated set of measures on S, $\mathfrak{M} \ll \lambda$; for any μ in $\mathfrak{M}$ write $f_\mu = d\mu/d\lambda$ and $K_\mu = \{x: f_\mu(x) > 0\}$. We define (for the purposes of this proof only) a *kernel* as a set K in S such that, for some measure μ in $\mathfrak{M}$, $K \subset K_\mu$ and $\mu(K) > 0$; we define a *chain* as a disjoint union of kernels. Since $\lambda(K) > 0$ for every kernel K, it follows from the finiteness of λ that every chain is a *countable* disjoint union of kernels. It follows also from these definitions that if C is a measurable subset of a chain, such that $\mu(C) > 0$ for at least one measure μ in $\mathfrak{M}$, then C is a chain, and that a disjoint union of chains is a chain. The last two remarks imply, through the usual process of disjointing any countable union, that a countable (but not necessarily disjoint) union of chains is a chain.

Let $\{C_j\}$ be a sequence of chains such that, as $j \to \infty$, $\lambda(C_j)$ approaches the supremum of the values of λ on chains. If $C = \bigcup_{j=1}^{\infty} C_j$, then C is a chain for which $\lambda(C)$ is maximal. The definition of a chain yields the existence of a sequence $\{K_i\}$ of kernels such that $C = \bigcup_{i=1}^{\infty} K_i$, and the definition of a kernel yields the existence, for each $i = 1, 2, \cdots$, of a measure μ_i in $\mathfrak{M}$ such that $K_i \subset K_{\mu_i}$ and $\mu_i(K_i) > 0$. We write $\mathfrak{N} = \{\mu_1, \mu_2, \cdots\}$; since $\mathfrak{N} \subset \mathfrak{M}$, the relation $\mathfrak{N} \ll \mathfrak{M}$ is trivial. We shall prove that $\mathfrak{M} \ll \mathfrak{N}$.

Suppose that $E \in S$, $\mu_i(E) = 0$ for $i = 1, 2, \cdots$, and let μ be any measure in $\mathfrak{M}$. It is to be proved that $\mu(E) = 0$. Since $\mu(E - K_\mu) = 0$, there is no loss of generality in assuming that $E \subset K_\mu$. If $\mu(E - C) > 0$, then $\lambda(E - C) > 0$ and therefore (since $E - C$ is a kernel) $E \cup C$ is a chain with $\lambda(E \cup C) > \lambda(C)$. Since this is impossible, it follows that $\mu(E - C) = 0$. Since $0 = \mu_i(E) = \mu_i(E \cap K_i) = \int_{E \cap K_i} f_{\mu_i}\, d\lambda$ and since $K_i \subset K_{\mu_i}$, it follows that $\lambda(E \cap K_i) = 0$. We conclude that $\lambda(E \cap C) = \sum_{i=1}^{\infty} \lambda(E \cap K_i) = 0$ and therefore $\mu(E \cap C) = 0$. Since $\mu(E) = \mu(E - C) + \mu(E \cap C)$, the proof of the lemma is complete.

5. Sufficient statistics for dominated sets. The statistic T is sufficient for a set $\mathfrak{M}$ of measures on S if, for every E in S, there exists a measurable function $p = p(E \mid y)$ on Y, such that

$$p_\mu(E \mid y) = p(E \mid y) \quad [\mu T^{-1}]$$

for every μ in $\mathfrak{M}$.[15] In other words, T is sufficient for $\mathfrak{M}$ if there exists a condi-

[15] The original definition of sufficiency was given by R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Roy. Soc. Phil. Trans.*, Series A, Vol. 222 (1922), pp. 309–368.

tional probability function common to every μ in $\mathfrak{M}$, or, crudely speaking, if the conditional distribution induced by T is independent of μ.

THEOREM 1. *A necessary and sufficient condition that the statistic T be sufficient for a dominated set $\mathfrak{M}$ of measures on S is that there exist a measure λ on S such that $\mathfrak{M} \equiv \lambda$ and such that $d\mu/d\lambda$ (ϵ) $T^{-1}(T)$ for every μ in $\mathfrak{M}$.*

Proof of necessity. Let $\mathfrak{N} = \{\mu_1, \mu_2, \cdots\}$ be a countable subset equivalent to $\mathfrak{M}$ (Lemma 7), and write λ for the measure on S defined by

$$\lambda(E) = \sum_{i=1}^{\infty} a_i \mu_i(E),$$

where $a_i = 1/2^i \mu_i(X)$, $i = 1, 2, \cdots$. Clearly $\mathfrak{M} \equiv \lambda$.

If p is a conditional probability function common to every μ in $\mathfrak{M}$, then, for every F in T,

$$\lambda(E \cap T^{-1}(F)) = \sum_{i=1}^{\infty} a_i \mu_i(E \cap T^{-1}(F))$$

$$= \sum_{i=1}^{\infty} a_i \int_F p(E \mid y)\, d\mu_i T^{-1}(y) = \int_F p(E \mid y)\, d\lambda T^{-1}(y),$$

i.e. p serves also as a conditional probability for λ.

Take any fixed μ in $\mathfrak{M}$, write $d\mu/d\lambda = f$, and $e_\lambda(f \mid y) = g(y)$; then $d\mu T^{-1} = g d\lambda T^{-1}$, and we have, for every E in S,

$$\int_E f(x)\, d\lambda(x) = \mu(E) = \int p(E \mid y)\, d\mu T^{-1}(y)$$

$$= \int p(E \mid y) g(y)\, d\lambda T^{-1}(y) = \int e_\lambda(\chi_E \mid y) e_\lambda(gT \mid y)\, d\lambda T^{-1}(y)$$

$$= \int e_\lambda(\chi_E \cdot gT \mid y)\, d\lambda T^{-1}(y) = \int \chi_E(x) gT(x)\, d\lambda(x) = \int_E gT(x)\, d\lambda(x).$$

The desired result, $f(x) = gT(x)$ $[\lambda]$, follows from a comparison of the first and last terms in the last written chain of equations.

Proof of Sufficiency. We shall prove that p_λ is a conditional probability function common to every μ in $\mathfrak{M}$. Take any fixed E in S and μ in $\mathfrak{M}$ and write $d\mu/d\lambda = gT$. If the measure ν is defined by $d\nu = \chi_E d\mu$, then $d\nu T^{-1} = p_\mu d\mu T^{-1}$, where $p_\mu = p_\mu(E \mid y)$. The hypothesis $d\mu = gT d\lambda$ implies that $d\mu T^{-1} = g d\lambda T^{-1}$ and hence that

$$d\nu T^{-1} = p_\mu \cdot g d\lambda T^{-1}.$$

On the other hand $d\nu = \chi_E d\mu = \chi_E \cdot gT d\lambda$, so that

$$d\nu T^{-1} = e_\lambda d\lambda T^{-1},$$

where $e_\lambda = e_\lambda(\chi_E \cdot gT \mid y) = p_\lambda(E \mid y) g(y)$. It follows from a comparison of the two expressions for $d\nu T^{-1}$ that

$$p_\mu(E \mid y) g(y) = p_\lambda(E \mid y) g(y)\ [\lambda T^{-1}].$$

Since the relation $d\mu T^{-1} = gd\lambda T^{-1}$ clearly implies that $g(y) \neq 0$ $[\mu T^{-1}]$ (i.e. that $\mu T^{-1}(\{y: g(y) = 0\}) = 0$), it follows, finally that

$$p_\mu(E \mid y) = p_\lambda(E \mid y) \; [\mu T^{-1}].$$

6. Special criteria for sufficiency. Theorem 1 may be recast in a form more akin in spirit to previous investigations of the concept of sufficiency.[16]

COROLLARY 1. *A necessary and sufficient condition that the statistic T be sufficient for a dominated set $\mathfrak{M}$ ($\ll \lambda_0$) of measures on S is that, for every μ in $\mathfrak{M}$, $f_\mu = d\mu/d\lambda_0$ be factorable in the form $f_\mu = g_\mu \cdot t$, where $0 \leqq g_\mu$ (ϵ) $T^{-1}(\mathfrak{T})$, $0 \leqq t$, t and $g_\mu \cdot t$ are integrable with respect to λ_0, and t vanishes $[\lambda_0]$ on each set in S for which every μ in $\mathfrak{M}$ vanishes.*

In more customary statistical language the condition asserts essentially that "each density is factorable into a function of the statistic alone and a function independent of the parameter."

PROOF. If T is sufficient for $\mathfrak{M}$, then there exists a measure λ with the properties described in Theorem 1. It follows that

$$f_\mu = \frac{d\mu}{d\lambda_0} = \frac{d\mu}{d\lambda}\frac{d\lambda}{d\lambda_0}$$

and we may write $g_\mu = d\mu/d\lambda$ and $t = d\lambda/d\lambda_0$. The only assertion that is not immediately obvious is the one concerning the vanishing of t. To prove it, suppose that $\mu(E) = 0$ for every μ in $\mathfrak{M}$; the fact that then

$$0 = \lambda(E) = \int_E t(x)\, d\lambda_0(x)$$

implies the desired conclusion.

If, conversely, $f_\mu = g_\mu \cdot t$, then we may write $d\lambda = td\lambda_0$. The relation $\mathfrak{M} \equiv \lambda$ follows from the statement concerning the vanishing of t, and the relation $d\mu/d\lambda$ (ϵ) $T^{-1}(\mathfrak{T})$ is implied by the equation $d\mu = g_\mu \cdot td\lambda_0 = g_\mu d\lambda$.

For the statement of the next consequence of Theorem 1 it is convenient to call a set $\mathfrak{M}$ of measures on S *homogencous* if $\mu \equiv \nu$ for every μ and ν in $\mathfrak{M}$.

COROLLARY 2. *A necessary and sufficient condition that the statistic T be sufficient for a homogeneous set $\mathfrak{M}$ of measures on S is that, for every μ and ν in $\mathfrak{M}$, $d\nu/d\mu$ (ϵ) $T^{-1}(\mathfrak{T})$.*

PROOF. Since a homogeneous set is dominated (by any one of its elements), Theorem 1 is applicable. If T is sufficient for $\mathfrak{M}$ and if λ has the properties described in Theorem 1, then $d\nu/d\mu = (d\nu/d\lambda)/(d\mu/d\lambda)$. The converse follows, through Theorem 1, by letting λ be any measure in $\mathfrak{M}$.

We shall say that the statistic T is *pairwise sufficient* for a set $\mathfrak{M}$ of measures

[16] See J. Neyman, "Su un teorema concernente le cosiddette statistiche sufficienti," *Inst. Ital. Atti. Giorn.*, Vol. 6 (1935), pp. 320–334. In this paper Neyman is somewhat restricted by his use of classical analytical methods, but he points out the possibility and desirability of extending his results to a much more general domain. For a recent presentation of the theory and further references to the literature cf. H Cramér, *Mathematical Methods of Statistics*, Princeton, 1946

on S if it is sufficient for every pair $\{\mu, \nu\}$ of measures in $\mathfrak{M}$. In other words, T is pairwise sufficient for $\mathfrak{M}$ if, for every E in S and μ and ν in $\mathfrak{M}$, there exists a measurable function $p_{\mu\nu}(E \mid y)$ on Y such that

$$p_\mu(E \mid y) = p_{\mu\nu}(E \mid y)\ [\mu T^{-1}] \qquad \text{and} \qquad p_\nu(E \mid y) = p_{\mu\nu}(E \mid y)\ [\nu T^{-1}].$$

Since pairwise sufficiency is (at least apparently) weaker than sufficiency, it is not surprising that there is a simple criterion for it even in the case of quite arbitrary (not necessarily homogeneous or dominated) sets of measures.

COROLLARY 3. *A necessary and sufficient condition that T be pairwise sufficient for a set $\mathfrak{M}$ of measures on S is that, for any two measures μ and ν in $\mathfrak{M}$, $d\mu/d(\mu + \nu)$ (ϵ) $T^{-1}(T)$.*

PROOF. If T is sufficient for μ and ν, then there exists a measure $\lambda \equiv \mu + \nu$ such that $d\mu/d\lambda$ (ϵ) $T^{-1}(T)$ and $d\nu/d\lambda$ (ϵ) $T^{-1}(T)$. It follows that

$$\frac{d\mu}{d(\mu + \nu)} = \frac{d\mu}{d\lambda} \Big/ \frac{d(\mu + \nu)}{d\lambda} = \frac{d\mu}{d\lambda} \Big/ \left(\frac{d\mu}{d\lambda} + \frac{d\nu}{d\lambda}\right).$$

The sufficiency of the condition follows immediately by applying Theorem 1 to the two-element set $\{\mu, \nu\}$.

7. Pairwise sufficiency and likelihood ratios. It is sometimes convenient to express the result of Corollary 3 in slightly different language. If λ is a measure on S and if f and g are real valued measurable functions on X such that $\lambda(\{x: f(x) = g(x) = 0\}) = 0$, we shall say that the pair (f, g) is *admissible* $[\lambda]$. (Intuitively an admissible pair (f, g) is to be thought of as a ratio f/g, which, however, may not be formed directly at the points x for which $g(x) = 0$.) Two admissible pairs (f_1, g_1) and (f_2, g_2) will be called *equivalent* $[\lambda]$, in symbols $(f_1, g_1) \equiv (f_2, g_2)$ $[\lambda]$, if there exists a real valued measurable function t on X such that $t(x) \neq 0$ $[\lambda]$ and such that $f_1 = tf_2$ and $g_1 = tg_2$ $[\lambda]$. It is clear that the relation "$\equiv$ $[\lambda]$" is indeed an equivalence; the equivalence class containing the admissible pair (f, g) will be called the *ratio* of f and g and will be denoted by $f \mid g$. (A ratio may accordingly be described as a measurable function from X to the real projective line.) For a ratio $f \mid g$ we shall write $f \mid g$ (ϵ) $T^{-1}(T)$ $[\lambda]$ if the equivalence class $f \mid g$ contains a pair (f_0, g_0) which is admissible $[\lambda]$ and for which f_0 (ϵ) $T^{-1}(T)$ and g_0 (ϵ) $T^{-1}(T)$.

LEMMA 8. *If μ, ν, λ_1, and λ_2 are measures on S such that $\mu + \nu \ll \lambda_1$ and $\mu + \nu \ll \lambda_2$, then the pairs $(d\mu/d\lambda_1, d\nu/d\lambda_1)$ and $(d\mu/d\lambda_2, d\nu/d\lambda_2)$ are admissible $[\mu + \nu]$ and equivalent $[\mu + \nu]$.*

PROOF. The admissibility of, for instance, $(d\mu/d\lambda_1, d\nu/d\lambda_1)$ follows from the fact that $d\mu/d\lambda_1 \neq 0$ $[\mu]$ and $d\nu/d\lambda_1 \neq 0$ $[\nu]$, whence

$$(\mu + \nu)\left(\left\{x: \frac{d\mu}{d\lambda_1}(x) = \frac{d\nu}{d\lambda_1}(x) = 0\right\}\right) = 0.$$

To prove equivalence, we write $\lambda_1 + \lambda_2 = \lambda$. Since

$$\frac{d\mu}{d\lambda_1}\frac{d\lambda_1}{d\lambda} = \frac{d\mu}{d\lambda} = \frac{d\mu}{d\lambda_2}\frac{d\lambda_2}{d\lambda}, \qquad \frac{d\nu}{d\lambda_1}\frac{d\lambda_1}{d\lambda} = \frac{d\nu}{d\lambda} = \frac{d\nu}{d\lambda_2}\frac{d\lambda_2}{d\lambda},$$

since also $d\lambda_1/d\lambda \neq 0$ $[\lambda_1]$ and therefore $d\lambda_1/d\lambda \neq 0$ $[\mu + \nu]$, and since, similarly, $d\lambda_2/d\lambda \neq 0$ $[\mu + \nu]$, the conditions of the definition of equivalence are satisfied by $t = (d\lambda_2/d\lambda)/(d\lambda_1/d\lambda)$

If μ and ν are any two measures on $\mathcal{S}$ and if λ is any measure on $\mathcal{S}$ such that $\mu + \nu \ll \lambda$ (for instance if $\lambda = \mu + \nu$), then the ratio $d\mu/d\lambda \mid d\nu/d\lambda$, which according to Lemma 8 exists $[\mu + \nu]$ and is independent of λ, will be called the *likelihood ratio* of μ and ν and will be denoted by $d\mu \mid d\nu$. The result of Corollary 3 may be expressed in terms of likelihood ratios as follows.

THEOREM 2. *A necessary and sufficient condition that T be pairwise sufficient for a set $\mathfrak{M}$ of measures on $\mathcal{S}$ is that, for any two measures μ and ν in $\mathfrak{M}$, $d\mu \mid d\nu$ (ϵ) $T^{-1}(\boldsymbol{T})$.*

PROOF. If T is sufficient for μ and ν, then, by Corollary 3, $d\mu/d(\mu + \nu)$ (ϵ) $T^{-1}(\boldsymbol{T})$, $d\nu/d(\mu + \nu)$ (ϵ) $T^{-1}(\boldsymbol{T})$, and, by Lemma 8, $(d\mu/d(\mu + \nu), d\nu/d(\mu + \nu))$ is an admissible pair belonging to the equivalence class $d\mu \mid d\nu$. Suppose conversely that $f = d\mu/d(\mu + \nu)$, $g = d\nu/d(\mu + \nu)$, and let the real valued measurable functions t, f_0, and g_0 be such that $t \neq 0$ $[\mu + \nu]$, f_0 (ϵ) $T^{-1}(\boldsymbol{T})$, g_0 (ϵ) $T^{-1}(\boldsymbol{T})$, (f_0, g_0) is admissible $[\mu + \nu]$, and

$$f = t \cdot f_0, \qquad g = t \cdot g_0 \; [\mu + \nu].$$

Since f and g are non negative, it follows that $f = |t| \cdot |f_0|$ and $g = |t| \cdot |g_0|$ $[\mu + \nu]$, i.e. that there is no loss of generality in assuming that t, f_0, and g_0 are non negative. The relation $f + g = 1$ $[\mu + \nu]$ implies that $t \cdot (f_0 + g_0) = 1$ $[\mu + \nu]$; the fact that (f_0, g_0) is admissible $[\mu + \nu]$ then yields $t \epsilon T^{-1}(\boldsymbol{T})$. The proof is completed by comparing this result with the expressions for f and g in terms of f_0 and g_0 and applying Corollary 3.

8. Pairwise sufficiency versus sufficiency. In order to show that our results on pairwise sufficiency (in the preceding section and in the sequel) are not vacuous, we proceed now to exhibit a statistic which is, for a suitable set of measures, pairwise sufficient but not sufficient.

Let $X = \{(x, i): 0 \leqq x \leqq 1, i = 0, 1\}$ be the union of two unit intervals and let $Y = \{y: 0 \leqq y \leqq 1\}$ be a unit interval. In accordance with our basic convention, measurability in both X and Y is to be taken in the sense of Borel. The statistic T is defined by $T(x, i) = x$.

Write $X_0 = \{(x, 0): 0 \leqq x \leqq 1\}$ and $X_1 = \{(x, 1): 0 \leqq x \leqq 1\}$. Let μ be (linear) Lebesgue measure on the class $\mathcal{S}$ of Borel subsets of X, and define, whenever $E \epsilon \mathcal{S}$ and $0 \leqq \alpha \leqq 1$,

$$\mu_\alpha(E) = \tfrac{1}{2}[\mu(E \cap X_0) + \chi_{E \cap X_1}(\alpha, 1)].$$

Let ν be (linear) Lebesgue measure on the class $\mathcal{T}$ of Borel subsets of Y, and define, whenever $F \epsilon \mathcal{T}$ and $0 \leqq \alpha \leqq 1$,

$$\nu_\alpha(F) = \tfrac{1}{2}[\nu(F) + \chi_F(\alpha)].$$

Clearly $\nu_\alpha = \mu_\alpha T^{-1}$; we write $\mathfrak{M} = \{\mu_\alpha : 0 \leqq \alpha \leqq 1\}$.

If $\delta(y, \alpha)$ is defined to be 1 or 0 according as $y = \alpha$ or $y \neq \alpha$, if $\delta'(y, \alpha) = 1 - \delta(y, \alpha)$, and if

$$p_\alpha(E \mid y) = \delta'(y, \alpha)\chi_E(y, 0) + \delta(y, \alpha)\chi_E(y, 1),$$

then a straightforward computation shows that

$$\mu_\alpha(E \cap T^{-1}(F)) = \int_F p_\alpha(E \mid y)\, d\nu_\alpha(y),$$

so that $p_\alpha(E \mid y) = p_{\mu_\alpha}(E \mid y)$ $[\nu_\alpha]$.

It is now easy to verify that T is pairwise sufficient for $\mathfrak{M}$. Indeed if α and β are any two different numbers in the closed unit interval, we may write

$$p(E \mid y) = \delta'(y, \alpha)\delta'(y, \beta)\chi_E(y, 0) + [\delta(y, \alpha) + \delta(y, \beta)]\chi_E(y, 1).$$

Since $\{y: p(E \mid y) \neq p_\alpha(E \mid y)\} = \{\beta\}$ and $\{y: p(E \mid y) \neq p_\beta(E \mid y)\} = \{\alpha\}$, it follows that $p(E \mid y) = p_\alpha(E \mid y)$ $[\nu_\alpha]$ and $p(E \mid y) = p_\beta(E \mid y)$ $[\nu_\beta]$.

To prove that T is not sufficient for $\mathfrak{M}$ we observe that $p_\alpha(X_1 \mid y) = \delta(y, \alpha)\chi_{X_1}(y, 1) = \delta(y, \alpha)$ and therefore

$$p_{\mu_\alpha}(X_1 \mid y) = \delta(y, \alpha)\ [\nu_\alpha].$$

Suppose that there is a conditional probability function p such that $p(E \mid y) = p_{\mu_\alpha}(E \mid y)$ $[\nu_\alpha]$. Then, in particular,

$$p(X_1 \mid y) = \delta(y, \alpha)\ [\nu_\alpha].$$

Since $\nu_\alpha(\{\alpha\}) = \frac{1}{2} > 0$, it follows that

$$p(X_1 \mid \alpha) = \delta(\alpha, \alpha) = 1,$$

or, changing to a more suggestive notation, that $p(X_1 \mid y) = 1$ for all y. We have, however,

$$\begin{aligned}\nu_\alpha(\{y: p_\alpha(X_1 \mid y) = 0\}) &= \nu_\alpha(\{y: \delta(y, \alpha) = 0\}) \\ &= \nu_\alpha(\{y: y \neq \alpha\}) = \tfrac{1}{2},\end{aligned}$$

so that $\nu_\alpha(\{y: p_{\mu_\alpha}(X_1 \mid y) = 0\}) = \frac{1}{2}$. This contradiction shows the impossibility of the existence of a conditional probability function common to every μ in $\mathfrak{M}$.

This example shows also that, in a sense, sufficiency is more fundamental than pairwise sufficiency. If, for instance, we imagine that it is important to a statistician that he either estimate α sharply or refrain from estimating it altogether, then he is by no means as well off with the observation of y as with that of x.

9. Pairwise sufficiency for dominated sets. We now proceed to show that for dominated sets of measures no such example as the one in the preceding section exists, or, in other words, that for dominated sets the concepts of pairwise sufficiency and sufficiency do coincide.

LEMMA 9. *If T is pairwise sufficient for a set $\{\mu_0, \mu_1, \mu_2\}$ of three measures on $\mathcal{S}$, then*[17]

$$\frac{d\mu_0}{d(\mu_0 + \mu_1 + \mu_2)} \ (\epsilon)T^{-1}(T).$$

PROOF. According to Corollary 3,

$$f_1 = \frac{d\mu_0}{d(\mu_0 + \mu_1)} \ (\epsilon)T^{-1}(T) \quad \text{and} \quad f_2 = \frac{d\mu_0}{d(\mu_0 + \mu_2)} \ (\epsilon)T^{-1}(T).$$

Since $d\mu_0 = f_1 d(\mu_0 + \mu_1) = f_2 d(\mu_0 + \mu_2)$, we have $f_1 d\mu_0 = f_1 f_2 d(\mu_0 + \mu_2)$ and $f_2 d\mu_0 = f_1 f_2 d(\mu_0 + \mu_1)$, so that

$$(f_1 + f_2 - f_1 f_2) d\mu_0 = f_1 f_2 d(\mu_0 + \mu_1 + \mu_2).$$

If we write $d\mu_0 = f d(\mu_0 + \mu_1 + \mu_2)$, then it follows that

$$(f_1 + f_2 - f_1 f_2) f = f_1 f_2 \ [\mu_0 + \mu_1 + \mu_2].$$

Since $0 \leqq f_1 \leqq 1$ and $0 \leqq f_2 \leqq 1$, the equation $f_1 + f_2 - f_1 f_2 = 0$ is equivalent to $f_1 = f_2 = 0$. Since $\mu_0(\{x: f_1(x) = f_2(x) = 0\}) = 0$, it follows that f may be redefined, if necessary, to be 0 on the set $\{x: f_1(x) = f_2(x) = 0\}$ without affecting the relation $d\mu_0 = f d(\mu_0 + \mu_1 + \mu_2)$, since outside this set $f = f_1 f_2 / (f_1 + f_2 - f_1 f_2)$, the proof of the lemma is complete.

LEMMA 10. *If T is pairwise sufficient for a finite set $\{\mu_0, \mu_1, \cdots, \mu_k\}$ of measures on $\mathcal{S}$, then $d\mu_0 / d(\sum_{i=0}^{k} \mu_i)$ (ϵ) $T^{-1}(T)$.*

PROOF. For $k = 1$ the conclusion is a restatement of the hypothesis; we proceed by induction. Given $\mu_0, \mu_1, \cdots, \mu_{k+1}$, we write $\mu = \sum_{i=1}^{k} \mu_i$. Then $d\mu_0 / d(\mu_0 + \mu)$ (ϵ) $T^{-1}(T)$ by the induction hypothesis and $d\mu_0 / d(\mu_0 + \mu_{k+1})$ (ϵ) $T^{-1}(T)$ by Corollary 3. Lemma 9 may then be applied to $\{\mu_0, \mu, \mu_{k+1}\}$ and yields the desired conclusion.

LEMMA 11. *If $\{\mu_0, \mu_1, \mu_2 \cdots\}$ is a sequence of measures on $\mathcal{S}$ such that $\sum_{i=0}^{\infty} \mu_i(X) < \infty$, if, for every E in $\mathcal{S}$, $\mu(E) = \sum_{i=0}^{\infty} \mu_i(E)$; and if λ is a measure $\mathcal{S}$ such that $\mu_i \ll \lambda$ for $i = 0, 1, 2, \cdots$, then*

$$\lim_k d(\textstyle\sum_{i=0}^{k} \mu_i)/d\lambda = d\mu/d\lambda \ [\lambda].$$

PROOF. Since $0 \leqq d(\sum_{i=0}^{k} \mu_i)/d\lambda = \sum_{i=0}^{k} (d\mu_i/d\lambda) \leq d\mu/d\lambda \ [\lambda]$, the series $\sum_{i=0}^{\infty} (d\mu_i/d\lambda)$ does indeed converge to a measurable function f $[\lambda]$. Since, for every E in $\mathcal{S}$,

$$\int_E f \, d\lambda = \sum_{i=0}^{\infty} \int_E \frac{d\mu_i}{d\lambda} \, d\lambda = \sum_{i=0}^{\infty} \mu_i(E) = \mu(E),$$

we have $f = d\mu/d\lambda$ $[\lambda]$, as stated.

[17] In view of Theorem 1, Lemma 9 asserts that if T is pairwise sufficient for a set $\mathfrak{M}$ of three elements, then T is sufficient for $\mathfrak{M}$. Lemmas 10 and 12 extend this result to finite and countably infinite sets $\mathfrak{M}$ respectively. Since every countable set of measures is dominated, the final result, Theorem 3, contains all these preliminaries as special cases.

LEMMA 12. *If* $\{\mu_0, \mu_1, \mu_2, \cdots\}$ *is a sequence of measures on* S *such that* $\sum_{i=0}^{\infty} \mu_i(X) < \infty$, *and if, for every* E *in* S, $\mu(E) = \sum_{i=0}^{\infty} \mu_i(E)$, *then*

$$\lim_k d\mu_0/d(\textstyle\sum_{i=0}^{k} \mu_i) = d\mu_0/d\mu \ [\mu].$$

If, in addition, T *is pairwise sufficient for the sequence* $\{\mu_0, \mu_1, \mu_2 \cdots\}$, *then* $d\mu_0/d\mu$ (ϵ) $T^{-1}(\mathcal{T})$.

PROOF. We have, for $k = 0, 1, 2, \cdots$,

$$\frac{d\mu_0}{d(\sum_{i=0}^{k} \mu_i)} \cdot \frac{d(\sum_{i=0}^{k} \mu_i)}{d\mu} = \frac{d\mu_0}{d\mu}.$$

If we write $\lambda = \mu$, then the hypotheses of Lemma 11 are satisfied and, consequently, the second factor on the left side converges to 1 $[\mu]$; it follows that the first factor converges to $d\mu_0/d\mu$ $[\mu]$. The second assertion of the lemma follows from Lemma 10.

THEOREM 3. *A necessary and sufficient condition that* T *be sufficient for a dominated set* $\mathfrak{M}$ *of measures on* S *is that* T *be pairwise sufficient for* $\mathfrak{M}$.

PROOF. The necessity of the condition is obvious. To prove its sufficiency, let $\mathfrak{N} = \{\mu_1, \mu_2, \cdots\}$ be a countable subset of $\mathfrak{M}$ which is equivalent to $\mathfrak{M}$ (Lemma 7), and let μ_0 be an arbitrary measure in $\mathfrak{M}$. Since the sufficiency or pairwise sufficiency of T remains unaltered if some or all of the measures in $\mathfrak{M}$ are replaced by positive constant multiples of themselves, we may assume that $\sum_{i=0}^{\infty} \mu_i(X) < \infty$. If we write, for every E in S, $\lambda(E) = \sum_{i=1}^{\infty} \mu_i(E)$, then the pairwise sufficiency of T and Lemma 12 imply that $d\mu_0/d(\mu_0 + \lambda)$ (ϵ) $T^{-1}(\mathcal{T})$. The relation

$$\frac{d\mu_0}{d\lambda} = \frac{d\mu_0}{d(\mu_0 + \lambda)} \cdot \frac{d(\mu_0 + \lambda)}{d\lambda} = \frac{d\mu_0}{d(\mu_0 + \lambda)} \cdot \left(\frac{d\lambda}{d(\mu_0 + \lambda)}\right)^{-1}$$
$$= \frac{d\mu_0}{d(\mu_0 + \lambda)} \left(1 - \frac{d\mu_0}{d(\mu_0 + \lambda)}\right)^{-1}$$

implies that $d\mu_0/d\lambda$ (ϵ) $T^{-1}(\mathcal{T})$; an application of Theorem 1 concludes the proof.

A comparison of Theorems 1 and 2 and Corollary 3 yields immediately the following consequence of Theorem 3.

COROLLARY 4. *A necessary and sufficient condition that the statistic* T *be sufficient for a dominated set* $\mathfrak{M}$ *of measures on* S *is that, for any two measures* μ *and* ν *in* $\mathfrak{M}$, $d\mu/d(\mu + \nu)$ (ϵ) $T^{-1}(\mathcal{T})$, *or, equivalently,* $d\mu \mid d\nu$ (ϵ) $T^{-1}(\mathcal{T})$.

10. The value of sufficient statistics in statistical methodology. We gather from conversations with some able and prominent mathematical statisticians that there is doubt and disagreement about just what a sufficient statistic is sufficient to do, and in particular about in what sense if any it contains "all the information in a sample." We therefore conclude this paper with a brief explanation of a point of view which, while not original with us, has not received due publicity.

Suppose a statistician $\mathcal{S}$ is to be shown an observation x drawn at random from some sample space (X, S) on which an unknown measure, μ, of a set $\mathfrak{M}$ of possible measures obtains, while for the same observation x another statistician $\mathcal{T}$ is only to be shown the value $T(x)$ of some statistic T sufficient for $\mathfrak{M}$. It is clear that $\mathcal{S}$ is as well off as $\mathcal{T}$; we shall argue that $\mathcal{T}$ is also as well off as $\mathcal{S}$.

Suppose $\mathcal{S}$ has decided how to use his datum, that, in other words, he has decided just what he will do (or, in particular, say) in the event of each possible x. His program can then be described schematically by saying that he has selected some function f (of the points x) which, without serious loss of generality, may be supposed to take real values. Now $\mathcal{S}$'s only real concern is for the probability distribution of f given μ, i.e. for the function φ of a real variable c, defined by

$$\varphi(c) = \mu(\{x: f(x) < c\}) = \mu(E(c)).$$

But $\mathcal{T}$ can if he wishes achieve exactly the same results as $\mathcal{S}$, in the following way. Let him, on learning the value of $T(x)$, select a real number f, with the aid of a "random machine" which produces numerical values according to the known distribution function ψ, defined by

$$\psi(c) = p(E(c) \mid T(x)).$$

Then, for any μ in $\mathfrak{M}$, the probability that $\mathcal{T}$ will select a value less than c is

$$\int p(E(c) \mid y) d\mu T^{-1}(y) = \mu(E(c)) = \varphi(c).$$

Thus $\mathcal{T}$ is at no disadvantage, save for the mechanical one of having to manipulate a random machine, and he may fairly be said to have as much information as $\mathcal{S}$.

As a matter of fact we know of no practical situation in which $\mathcal{T}$ would actually go to the trouble of using a random machine. There are some situations in which he should in principle do so, but in which practical statisticians have not, so far as we know, thought it worth while. If, for example, an outcome consists of a sequence of n heads and tails resulting from n spins of a coin the heads ratio of which is known to be either one half or one quarter, then a sufficient statistic is the number of heads which occur in the sequence. In basing a decision on the outcome of this program both $\mathcal{S}$ and, to a still greater extent, $\mathcal{T}$ have (according to Wald's theory of minimum risk) something to gain by recourse to a random machine. There are, on the other hand, many technical desiderata which sufficient statistics meet exactly without recourse to random machines. Thus, as Blackwell has shown,[18] if $\mathcal{S}$ has an unbiased estimate, R, of some parameter, $\mathcal{T}$ can find a function R^*, defined by $R^*(y) = e(R \mid y)$, which is an unbiased estimate of that parameter, with variance not greater than that of R. More generally, if R is *any* estimate with finite mean square deviation from a parameter, then it is easy to show with Blackwell's methods that R^*

[18] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math Stat.*, Vol. 18 (1947), pp 105–110.

has no larger a mean square deviation than R. Finally it is a well known fact that, under suitable hypotheses, if there exists a maximum likelihood estimate R of some parameter, then R depends only on y.

We think that confusion has from time to time been thrown on the subject by (a) the unfortunate use of the term "sufficient estimate," (b) the undue emphasis on the factorability of sufficient statistics, and (c) the assumption that a sufficient statistic contains all the information in only the technical sense of "information" as measured by variance.

ON DESIGNING SINGLE SAMPLING INSPECTION PLANS

By Frank E. Grubbs

Ballistic Research Laboratories, Aberdeen Proving Ground, Md.

1. Summary. In designing single sampling inspection plans, a problem is to find the acceptance number, c, and the smallest sample size, n, such that if the fraction defective of the material inspected is equal to an acceptable value, p_1, a large percentage, say, 95% of such lots will be accepted under the sample criteria, whereas if the fraction defective of the material inspected is objectionable and equal to p_2 (where $p_1 < p_2$), then a large percentage, say, 90% of such lots will be rejected. A solution to this problem for the case where the lot size is large compared to the sample size is given in this paper and tables are provided for quick determination of the sample size n and acceptance number c.

2. Introduction. In sampling inspection of material one practice is to set an acceptable quality level $= p_1$, say, such that the consumer desires to accept practically all—95% or more—of lots of fraction defective p_1 or less (and hence desires to reject at most a maximum of about 5% of lots which are of quality p_1 or better) and to set also an objectionable fraction defective $= p_2$, say, which represents quality so poor that the consumer cannot afford to accept more than about 10% or less of lots of this quality or poorer.[1] From the standpoint of the producer, he should have very few rejections, 5% or less, for his submitted lots the fractions defective of which are equal to or better (less) than p_1, whereas he should be willing and also expect to suffer increasingly more rejections if his process average percent defective departs from the acceptable quality level p_1 toward poor or objectionable quality. In this connection, if we are given p_1 an acceptable quality level, p_2 an objectionable percent defective, the risk $\alpha = 5\%$ of rejecting a lot of fraction defective p_1, and the risk $\beta = 10\%$ of accepting a lot of the objectionable fraction defective p_2, a problem of importance in single sampling inspection is to find the smallest sample size n and the acceptance number c which will approximate closely the protection stated above. Due to the discrete nature of n and c, it is not usually possible to find n and c such that precisely the above protection is guaranteed; however, it is possible to pick that single sampling plan which, for all practical purposes, gives the desired protection, i.e. it is possible to select that single sampling plan which more nearly satisfies

[1] When this paper was first presented for publication, the percent defectives p_1 and p_2 were labeled "Acceptable Quality Level" and "Lot Tolerance Percent Defective," respectively. In view of the suggestions of H. G. Romig and H. F. Dodge, strict reference to these particular terms have been avoided in order that the percent defectives p_1 and p_2 would appear in a more generalized form. This recommendation is considered especially desirable in view of the fact that Table I and Table II of the paper are percentage points of the Binomial Distribution and hence are useful in problems other than that of designing single sampling inspection plans.

the above protection requirements than any other plan. The values of n and c can be found simply by looking for an entry in Table I below which is close to p_1 and an entry in Table II close to p_2 such that column heading c and row heading n in Table I correspond exactly with the respective column and row headings in Table II. For the sample sizes n, acceptance numbers c and quality levels p covered in Tables I and II, the above procedure makes unnecessary any computation of or any approximation to the sample size and acceptance number. It will be noticed, however, that usually the proper choice of c is clear whereas some slight judgment may be necessary in selecting n.

It is remarked also that Tables I and II solve the equivalent problem of finding n and c in connection with testing the hypothesis H_0 that the fraction defective of the Binomial population sampled is p_1 or less as against an alternative hypothesis H_1 which states that the fraction defective of the lot, population, process, etc., sampled is p_2 or greater ($p_2 > p_1$), where $\alpha = .05$ is the maximum risk of erroneously rejecting H_0 when it is true and $\beta = .10$ is the maximum risk of erroneously accepting H_0 when the alternative H_1 is true.

The solution to the problem of finding an appropriate single sampling plan in this paper is given by solving the infinite case, i.e. by assuming the lot to be an infinite Binomial population. In practice lots are of finite size. However, it is well known that Binomial probabilities (infinite universe) give excellent practical approximations to Hypergeometric probabilities (finite lot) provided the sample size is only a small percentage of the lot size. Hence, the reader is warned in using the tables for sampling inspection problems that the lot size should be at least 10 or 15 times the sample size.

3. Basis for construction of Table I and Table II. It is well known that if $P(c, n, p)$ represents the probability of obtaining c or less defectives in a random sample of size n from a Binomial Population of fraction defective p, then the relation between $P(c, n, p)$ and the Incomplete Beta Function Ratio is given by

$$(1) \quad P(c, n, p) = I_{1-p}(n - c, c + 1) = \frac{1}{\beta(n - c, c + 1)} \int_0^{1-p} x^{n-c-1}(1 - x)^c \, dx.$$

Consequently, using a table of percentage points for the Incomplete Beta Function (1), values of p_1 can be found for Table I such that

$$P(c, n, p_1) = .95,$$

and values of p_2 can be found for Table II presented at the end such that

$$P(c, n, p_2) = .10.$$

Also, Table I and Table II can be computed by using percentage points of the F-distribution (2). Upon making the transformation

$$x = \frac{2(n - c)}{2(n - c) + 2(c + 1)F}$$

in (1) above to the F-distribution, we obtain easily that

$$(2) \quad P(c, n, p) = \frac{1}{\beta(c+1, n-c)} \int_{(n-c)p/(c+1)q}^{\infty} [2(c+1)]^{c+1} [2(n-c)]^{n-c} F^{c} \cdot [2(n-c) + 2(c+1)F]^{-n-1} \, dF,$$

where $q = 1 - p$.

With the aid of a table of percentage points of the F-distribution (2), we may determine for various combinations of $n - c$ and $c + 1$ those values of p such that

$$P(c, n, p_1) = .95 \quad \text{for Table I;}$$

and

$$P(c, n, p_2) = .10 \quad \text{for Table II.}$$

In fact, if $P(c, n, p) = \alpha$, then

$$\frac{(n-c)p}{(c+1)q} = F_{\alpha}\{2(c+1), 2(n-c)\},$$

or

$$p = \frac{(c+1)F_{\alpha}\{2(c+1), 2(n-c)\}}{(n-c) + (c+1)F_{\alpha}\{2(c+1), 2(n-c)\}},$$

for which relation values of p_1 for $\alpha = .95$ are given in Table I below and values of p_2 for $\alpha = .10$ are given in Table II.

Although the 95% points are not given directly in (2), they are easily obtainable from the relation

$$F_{.95}(\nu_1, \nu_2) = \frac{1}{F_{.05}(\nu_2, \nu_1)}.$$

Interpolation was required for the great majority of the entries in Tables I and II. The values given were obtained by harmonic or linear interpolation using References [1] and [2] and are believed accurate to within one unit in the last place.

It will be noticed that if the chosen acceptable quality level, p_1, is greater than the appropriate tabulated value in Table I for the single sampling plan (n, c), then the operating characteristic curve will pass below the point $(p_1, .95)$. That is, the risk of rejection under the sampling plan for lots of fraction defective p_1 will be somewhat more than 5%. On the other hand, if a selected acceptable quality level p_1 is less than the appropriate entry in Table I, the risk of rejection for a product of fraction defective p_1 will be less than 5%. Similar considerations apply also to the fractions defective, p_2, in Table II.

4. Single sampling plans based on the Poisson approximation to the binomial. Tables I and II are useful for determination of a single sampling plan when the

desired percent defectives are listed and n does not exceed 150. Table III is particularly useful in designing a single sampling plan when we are interested in fractions defective not greater than about .10. A somewhat similar procedure has already been suggested by Peach and Littauer [3]. If we designate by $P(c, a)$ the sum of individual Poisson probabilities,

$$P(c, a) = \sum_{m=0}^{c} \frac{e^{-a} a^m}{m!},$$

then Table III gives values $a_1 = np_1$ of a for which

$$P(c, a_1) = .95$$

and values $a_2 = np_2$ of a for which

$$P(c, a_2) = .10.$$

Hence, to find the single sampling plan whose operating characteristic curve passes nearly through the points $(p_1, .95)$ and $(p_2, .10)$ one merely divides values of a_1 in Table III for various values of c by the acceptable quality level p_1 and divides values of a_2 in Table III by the objectionable percent defective p_2. Then the acceptance number c is picked for which a_1/p_1 most nearly equals a_2/p_2 and the approximate sample size n may be determined by rounding to an integer the average of the two approximately equal numbers a_1/p_1 and a_2/p_2.

5. Example on the use of Tables I, II, III. Given an acceptable percent defective or quality level of .01 and an objectionable quality level of .10, it is desired to find the single sampling plan which will accept 95% of product which is of quality $p_1 = .01$ and which will reject 90% (or accept only 10%) of product of quality $p_2 = .10$. Looking in Table I for entries p_1 which approximately equal .01 and in Table II for entries p_2 which approximately equal .10 such that the c and n of Tables I and II correspond, we see that c must be equal to 1 whereas n may take possibly any one of the values 35, 36, 37, 38. In this connection, we have to set up some criteria for the choice of n. Although any of several criteria may be used, a reasonable criterion appears to involve picking n such that the sum of the absolute departures of the Operating Characteristic Curve from the risks $\alpha = .05$ at p_1 and $\beta = .10$ at p_2 is a minimum. This may be determined by using appropriate tables of Binomial Probabilities or by computing at p_1 and p_2 the chance of obtaining c or less defectives in n for the various possible combinations of c and n. If the above criterion were applied to the present example, the combination $c = 1$ and $n = 37$ would be selected, i.e. the single sampling plan would be $c = 1, n = 37$. For this sampling plan, the probability of passing at $p_1 = .01$ is .9471 and the probability of passing at $p_2 = .10$ is .1036. For the sake of expediency, another proposal would be merely to select somewhat of a "middle" value of n especially when the variation in sample size is slight.

If we use Table III for the above example, we can select n and c with the aid

of the following simple tabulation:

n	c			
	0	1	2	3
a_1/p_1	5.1	35.5	81.8	136.6
a_2/p_2	23.0	38.9	53.2	66.8

Since the sample sizes "cross" at $c = 1$, we would select $c = 1$ and $n = 1/2$ $(35.5 + 38.9) = 37.2$ or $n = 37$.

A use of Table I of some practical importance is in determining at a glance those values of p for which the probability of obtaining c or less defectives in a sample of n is equal to .95. As a matter of fact, a series of tables similar to Table I and Table II for which $P(c, n, p) = .99, .95, .90, .10, .05, .01$ etc. would be of considerable practical use.

Acknowledgment. The author is indebted to Miss Helen J. Coon for carrying out the computations for the tables.

TABLE I

Values of $p = p_1$ *such that* $P(c, n, p_1) = .95$

n	c										n
	0	1	2	3	4	5	6	7	8	9	
1	.0500										1
2	.0253	.224									2
3	.0170	.135	.368								3
4	.0127	.0976	.249	.473							4
5	.0102	.0764	.189	.343	.549						5
6	.00851	.0628	.153	.271	.418	.607					6
7	.00730	.0534	.129	.225	.341	.479	.652				7
8	.00639	.0464	.111	.193	.289	.400	.529	.688			8
9	.00568	.0410	.0978	.169	.251	.345	.450	.571	.717		9
10	.00512	.0368	.0873	.150	.222	.304	.393	.493	.606	.741	10
11	.00465	.0333	.0788	.135	.200	.271	.350	.436	.530	.636	11
12	.00427	.0305	.0719	.123	.181	.245	.315	.391	.473	.562	12
13	.00394	.0281	.0660	.113	.166	.224	.287	.355	.427	.505	13
14	.00366	.0260	.0611	.104	.153	.206	.264	.325	.390	.460	14
15	.00341	.0242	.0568	.0967	.142	.191	.244	.300	.360	.423	15
16	.00320	.0227	.0531	.0903	.132	.178	.227	.279	.333	.391	16
17	.00301	.0213	.0499	.0846	.124	.166	.212	.260	.311	.364	17
18	.00285	.0201	.0470	.0797	.116	.156	.199	.244	.291	.341	18
19	.00270	.0190	.0445	.0753	.110	.147	.188	.230	.274	.320	19
20	.00256	.0181	.0422	.0714	.104	.140	.177	.217	.259	.302	20
21	.00244	.0172	.0401	.0678	.0988	.132	.168	.206	.245	.286	21
22	.00233	.0164	.0382	.0646	.0941	.126	.160	.196	.233	.271	22
23	.00223	.0157	.0365	.0617	.0898	.120	.152	.186	.222	.258	23
24	.00213	.0150	.0350	.0590	.0859	.115	.146	.178	.212	.246	24
25	.00205	.0144	.0335	.0566	.0823	.110	.139	.170	.202	.236	25
26	.00197	.0138	.0322	.0543	.0790	.106	.134	.163	.194	.226	26
27	.00190	.0133	.0310	.0522	.0759	.101	.129	.157	.186	.217	27
28	.00183	.0128	.0298	.0503	.0731	.0977	.124	.151	.179	.208	28
29	.00177	.0124	.0288	.0485	.0705	.0942	.119	.145	.172	.200	29
30	.00171	.0120	.0278	.0469	.0681	.0909	.115	.140	.167	.193	30
31	.00165	.0116	.0269	.0453	.0658	.0878	.111	.135	.161	.187	31
32	.00160	.0112	.0260	.0438	.0637	.0850	.107	.131	.155	.180	32
33	.00155	.0109	.0252	.0425	.0617	.0823	.104	.127	.150	.175	33
34	.00151	.0106	.0245	.0412	.0598	.0798	.101	.123	.146	.169	34
35	.00146	.0102	.0238	.0400	.0580	.0774	.0978	.119	.141	.164	35

TABLE I—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
36	.00142	.00996	.0231	.0389	.0564	.0752	.0950	.116	.137	.159	36
37	.00139	.00969	.0225	0378	.0548	.0731	0923	.112	.133	.155	37
38	.00135	.00943	.0219	.0368	.0533	.0711	.0898	.109	.130	.150	38
39	.00131	.00919	.0213	.0358	.0519	.0692	.0874	.106	.126	.146	39
40	.00128	.00896	.0208	.0349	.0506	.0674	.0851	.104	123	.142	40
41	.00125	.00874	.0202	.0340	.0493	.0657	.0830	.101	.120	.139	41
42	.00122	.00853	.0198	.0332	.0481	.0641	.0809	.0985	.117	.135	42
43	.00119	.00833	.0193	.0324	0470	.0626	.0790	.0961	.114	.132	43
44	.00117	.00814	.0188	.0317	.0459	.0611	.0771	.0938	.111	.129	44
45	.00114	.00795	.0184	.0309	.0448	.0597	.0754	.0917	.109	.126	45
46	.00111	.00778	.0180	.0302	.0438	.0584	.0737	.0896	.106	.123	46
47	00109	.00761	.0176	.0296	.0429	.0571	.0720	.0876	.104	.120	47
48	.00107	.00745	.0172	.0290	.0420	.0559	.0705	.0857	.101	.118	48
49	.00105	.00730	.0169	.0284	.0411	.0547	.0690	.0839	.0993	.115	49
50	.00103	.00715	.0166	.0278	.0402	.0536	0676	.0822	.0972	.113	50
51	.00101	.00701	.0162	0272	.0394	.0525	.0662	.0805	.0953	.110	51
52	.000986	.00688	.0159	.0267	.0387	.0515	.0649	.0789	.0934	.108	52
53	.000967	.00675	.0156	.0262	.0379	.0505	.0637	.0774	.0916	.106	53
54	.000949	.00662	.0153	.0257	.0372	.0495	.0625	.0759	.0898	.104	54
55	.000932	.00650	.0150	.0252	.0365	.0486	.0613	.0745	.0881	.102	55
56	.000916	.00638	.0148	.0248	.0358	.0477	.0602	.0731	.0865	.100	56
57	.000899	00627	.0145	0243	.0352	.0468	.0591	.0718	.0849	.0984	57
58	.000884	.00616	0142	.0239	.0346	.0460	.0580	.0705	.0834	.0966	58
59	000869	.00606	.0140	.0235	.0340	.0452	.0570	.0693	.0820	.0949	59
60	.000855	.00595	.0138	.0231	0334	.0445	.0561	.0681	.0806	.0933	60
61	.000841	.00586	0135	.0227	.0329	.0437	.0551	.0670	.0792	.0917	61
62	.000827	.00576	.0133	.0223	.0323	.0430	.0542	.0659	.0779	.0902	62
63	.000814	.00567	.0131	.0220	.0318	.0423	.0533	.0648	.0766	.0887	63
64	.000801	.00558	.0129	.0216	.0313	.0416	.0525	.0637	.0754	.0873	64
65	.000789	.00549	.0127	.0213	.0308	.0410	.0516	.0627	.0742	.0859	65
66	.000777	00541	.0125	.0210	.0303	.0403	.0508	.0618	.0730	.0846	66
67	.000765	.00533	.0123	.0206	.0299	.0397	.0501	.0608	.0719	.0833	67
68	000754	00525	.0121	0203	0294	.0391	.0493	.0599	.0708	.0820	68
69	.000743	.00517	.0120	.0200	0290	.0385	.0486	.0590	.0698	.0808	69
70	000733	.00510	.0118	.0198	.0286	.0380	.0479	.0582	.0687	.0796	70

TABLE I—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
71	.000722	.00503	.0116	.0195	0282	.0374	.0472	.0573	.0678	.0785	71
72	.000712	.00496	.0115	.0192	.0278	.0369	.0465	.0565	.0668	.0773	72
73	.000702	.00489	.0113	.0189	.0274	.0364	.0459	0557	.0658	.0762	73
74	.000693	.00482	.0111	.0187	.0270	.0359	.0452	.0549	.0649	.0752	74
75	.000684	.00476	.0110	.0184	.0266	.0354	.0446	.0542	.0641	.0742	75
76	.000675	.00470	.0108	.0182	.0263	.0349	.0440	.0535	.0632	.0732	76
77	.000666	.00463	.0107	.0179	.0259	.0345	.0434	.0528	.0623	.0722	77
78	.000657	.00457	.0106	.0177	.0256	.0340	.0429	.0521	.0615	.0712	78
79	.000649	.00452	.0104	.0175	.0253	.0336	.0423	.0514	.0607	.0703	79
80	.000641	.00446	.0103	.0173	.0249	.0332	.0418	.0507	.0600	.0694	80
81	.000633	.00440	.0102	.0170	.0246	.0328	.0413	.0501	.0592	.0685	81
82	.000625	.00435	.0100	.0168	.0243	.0323	.0408	.0495	.0585	.0677	82
83	.000618	.00430	.00992	.0166	.0240	.0319	.0403	.0489	.0577	.0668	83
84	.000610	.00425	.00980	.0164	.0237	.0316	.0398	0483	.0570	.0660	84
85	.000603	.00420	.00969	.0162	.0235	.0312	.0393	.0477	.0564	.0652	85
86	.000596	.00415	.00957	.0160	.0232	.0308	.0388	.0471	.0557	.0645	86
87	.000589	.00410	.00946	.0159	.0229	.0305	.0384	.0466	.0550	.0637	87
88	.000583	.00405	.00936	.0157	.0227	.0301	.0379	.0460	.0544	.0630	88
89	.000576	.00401	.00925	.0155	.0224	.0298	.0375	0455	.0538	.0622	89
90	.000570	.00396	.00915	.0153	.0221	.0294	.0371	.0450	.0532	.0615	90
91	.000564	.00392	.00904	.0152	.0219	.0291	.0367	.0445	0526	.0608	91
92	.000557	.00388	.00895	.0150	.0217	.0288	.0363	.0440	.0520	.0602	92
93	.000551	.00383	.00885	.0148	.0214	.0285	.0359	.0435	.0514	.0595	93
94	.000546	.00379	.00875	.0147	.0212	.0282	.0355	.0431	.0509	0589	94
95	.000540	.00375	.00866	.0145	.0210	.0279	.0351	.0426	.0503	.0582	95
96	.000534	.00371	.00857	.0144	.0207	.0276	.0347	.0421	.0498	.0576	96
97	.000529	.00368	.00848	.0142	.0205	.0273	.0344	.0417	.0493	.0570	97
98	.000523	.00364	.00840	.0141	.0203	.0270	.0340	.0413	.0487	.0564	98
99	.000518	.00360	.00831	.0139	.0201	.0267	.0337	.0408	.0482	.0558	99
100	.000513	.00357	.00823	.0138	.0199	.0265	.0333	.0404	.0478	.0553	100
101	.000508	.00353	.00814	.0136	.0197	.0262	.0330	.0400	.0473	.0547	101
102	.000503	.00350	.00806	.0135	.0195	.0259	.0327	.0396	.0468	.0542	102
103	.000498	.00346	.00799	.0134	.0193	.0257	.0323	.0392	.0463	.0536	103
104	.000493	.00343	.00791	.0132	0191	.0254	.0320	.0389	.0459	.0531	104
105	.000488	.00339	.00783	.0131	.0189	.0252	.0317	.0385	.0454	.0526	105

TABLE I—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
106	.000484	.00336	.00776	.0130	.0188	0249	0314	.0381	.0450	.0521	106
107	.000479	.00333	.00768	0129	.0186	.0247	.0311	.0378	.0446	.0516	107
108	.000475	00330	00761	.0127	.0184	.0245	.0308	.0374	.0442	.0511	108
109	.000470	.00327	00754	.0126	.0182	.0242	.0305	.0370	.0438	.0506	109
110	000466	.00324	.00747	.0125	.0181	.0240	.0302	.0367	.0433	.0502	110
111	.000462	.00321	.00741	.0124	.0179	.0238	.0300	.0364	.0430	.0497	111
112	.000458	.00318	.00734	.0123	.0178	.0236	.0297	.0360	.0426	.0492	112
113	.000454	.00315	00727	.0122	.0176	.0234	.0294	.0357	.0422	.0488	113
114	.000450	.00313	.00721	.0121	.0174	.0232	.0292	.0354	.0418	.0484	114
115	.000446	.00310	.00715	.0120	.0173	.0230	.0289	.0351	.0414	.0479	115
116	.000442	.00307	.00709	.0119	.0171	.0228	.0287	.0348	.0411	.0475	116
117	.000438	.00305	.00702	.0118	.0170	.0226	.0284	.0345	.0407	.0471	117
118	.000435	.00302	.00696	.0117	.0168	.0224	.0282	.0342	.0404	.0467	118
119	.000431	.00299	.00691	.0116	.0167	.0222	.0279	.0339	.0400	.0463	119
120	.000427	.00297	.00685	.0115	.0166	.0220	.0277	.0336	.0397	.0459	120
121	.000424	.00294	.00679	.0114	.0164	.0218	.0275	.0333	.0394	.0455	121
122	.000420	.00292	.00674	0113	.0163	.0216	.0272	.0330	.0390	.0451	122
123	.000417	.00290	.00668	.0112	.0162	.0215	.0270	.0328	.0387	.0448	123
124	.000414	.00287	00663	.0111	.0160	.0213	.0268	.0325	.0384	.0444	124
125	.000410	.00285	.00657	.0110	.0159	.0211	0266	.0322	.0381	.0440	125
126	000407	.00283	.00652	.0109	.0158	0209	.0264	.0320	.0378	.0437	126
127	.000404	.00281	.00647	.0108	.0156	.0208	.0262	.0317	.0375	.0433	127
128	000401	.00278	.00642	.0107	.0155	0206	.0259	.0315	.0372	.0430	128
129	.000398	.00276	.00637	.0107	.0154	.0204	.0257	.0312	.0369	.0427	129
130	.000394	.00274	.00632	.0106	.0153	.0203	.0255	.0310	.0366	.0423	130
131	000391	.00272	.00627	.0105	.0152	.0201	.0253	.0308	.0363	.0420	131
132	.000389	.00270	.00622	.0104	.0150	.0200	.0252	.0305	.0360	.0417	132
133	.000386	.00268	.00618	.0103	.0149	.0198	.0250	.0303	.0358	.0414	133
134	.000383	.00266	.00613	.0103	.0148	.0197	.0248	.0301	.0355	.0410	134
135	.000380	.00264	.00608	.0102	.0147	.0195	.0246	.0298	.0352	.0407	135
136	.000377	.00262	.00604	.0101	.0146	.0194	.0244	.0296	.0350	.0404	136
137	.000374	.00260	.00599	.0100	.0145	.0192	.0242	.0294	.0347	.0401	137
138	.000372	.00258	.00595	.00996	.0144	.0191	.0240	.0292	.0344	.0398	138
139	.000369	.00256	.00591	.00989	.0143	.0190	.0239	.0290	.0342	.0395	139
140	.000366	.00254	.00587	.00982	.0142	.0188	.0237	.0288	.0339	.0393	140

TABLE I—*Concluded*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
141	.000364	.00253	.00582	.00975	.0141	.0187	.0235	.0285	.0337	.0390	141
142	.000361	.00251	.00578	.00968	.0140	.0186	.0234	.0283	.0335	.0387	142
143	.000359	.00249	.00574	.00961	.0139	.0184	.0232	.0281	.0332	.0384	143
144	.000356	.00247	.00570	.00954	.0138	.0183	.0230	.0279	.0330	.0382	144
145	.000354	.00246	.00566	.00948	.0137	.0182	.0229	.0278	.0328	.0379	145
146	.000351	.00244	.00562	.00941	.0136	.0180	.0227	.0276	.0325	.0376	146
147	.000349	.00242	.00559	.00935	.0135	.0179	.0226	.0274	.0323	.0374	147
148	.000346	.00241	.00555	.00928	.0134	.0178	.0224	.0272	.0321	.0371	148
149	.000344	.00239	.00551	.00922	.0133	.0177	.0223	.0270	.0319	.0369	149
150	.000342	.00237	.00547	.00916	.0132	.0176	.0221	.0268	.0317	.0366	150

TABLE II

Values of $p = p_2$ such that $P(c, n, p_2) = .10$

n	c										n
	0	1	2	3	4	5	6	7	8	9	
1	.900										1
2	.684	.949									2
3	.536	.804	.965								3
4	.438	.680	.857	.974							4
5	.369	.584	753	.888	.979						5
6	.319	.510	667	.799	.907	.983					6
7	280	.453	.596	.721	.830	.921	.985				7
8	.250	.406	.538	.655	.760	.853	.931	987			8
9	226	.368	.490	.599	.699	.790	.871	.939	.988		9
10	.206	337	.450	.552	.646	.733	.812	.884	.945	990	10
11	189	310	.415	.511	.599	.682	.759	.831	.895	.951	11
12	175	.288	386	.475	.559	.638	.712	.781	.846	904	12
13	.162	268	.360	.444	.523	.598	.669	.736	.799	.858	13
14	.152	.251	.337	.417	.492	.563	631	.695	.757	.815	14
15	.142	.236	317	.393	.464	.532	.596	.658	.718	.774	15
16	134	.222	300	.371	.439	.504	.565	.625	.682	.737	16
17	.127	.210	.284	.352	.416	478	.537	.594	650	.703	17
18	.120	.199	.269	.334	396	.455	.512	.567	.620	.671	18
19	.114	.190	257	.319	.378	.434	.489	.541	.592	.642	19
20	.109	.181	.245	.304	.361	415	.467	.518	.567	.615	20
21	104	.173	234	.291	.345	.397	.448	.497	.544	.590	21
22	.0994	.166	.224	.279	.331	381	.430	.477	.523	568	22
23	.0953	.159	.215	.268	.318	.366	.413	.459	.503	.546	23
24	.0915	.153	.207	.258	.306	.352	.398	.442	.485	.526	24
25	0880	.147	.199	.248	.295	.340	.383	.426	.467	.508	25
26	.0847	.142	.192	239	.284	.328	.370	.411	.451	.491	26
27	.0817	137	.185	.231	.275	.317	.358	.397	.436	.475	27
28	0789	132	.179	.223	.265	.306	.346	.385	.422	.459	28
29	.0763	.128	.173	.216	.257	.297	.335	.372	.409	.445	29
30	.0739	.124	.168	209	.249	.288	.325	.361	397	.432	30
31	0716	.120	.163	.203	.241	.279	.315	.350	.385	.419	31
32	0694	.116	.158	.197	.234	271	.306	.340	.374	.407	32
33	.0674	.113	.153	.191	.228	263	297	.331	.364	.396	33
34	0655	110	.149	.186	.221	.256	.289	322	.354	.385	34
35	0637	.107	.145	.181	216	.249	.282	.313	.345	.375	35

TABLE II—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
36	.0620	.104	.141	.176	.210	.242	.274	.305	.336	.366	36
37	.0603	.101	.138	.172	.205	.236	.267	.298	.327	.357	37
38	.0588	.0985	.134	.167	.199	.230	.261	.290	.319	.348	38
39	.0573	.0961	.131	.163	.195	.225	.254	.283	.312	.340	39
40	.0559	.0938	.128	.159	.190	.220	.248	.277	.305	.332	40
41	.0546	.0916	.125	.156	.186	.215	.242	.270	.298	.324	41
42	.0533	.0895	.122	.152	.181	.210	237	.264	.291	.317	42
43	.0521	.0875	.119	.149	.177	.205	.232	.259	.285	.310	43
44	.0510	.0856	.116	.146	.174	.201	.227	.253	.279	304	44
45	.0499	.0837	.114	.142	.170	.196	.222	.248	.273	.297	45
46	.0488	.0819	.112	.140	.166	.192	.218	.243	.268	.291	46
47	.0478	.0803	.109	.137	.163	.188	.213	.238	.262	.285	47
48	.0468	.0786	.107	.134	.160	.185	.209	.233	.257	.280	48
49	.0459	.0771	.105	.131	.157	.181	.205	.229	.252	.274	49
50	.0450	.0756	.103	.130	.154	.178	.201	.224	.248	.269	50
51	.0441	.0741	.101	.126	.151	.174	.197	.220	.243	.264	51
52	.0433	.0728	.0991	.124	.148	.171	.194	.216	.239	.259	52
53	.0425	.0714	.0973	.122	.145	.168	.190	.212	.235	.255	53
54	.0417	.0701	.0956	.120	.143	.165	.187	.208	.230	.250	54
55	.0410	.0689	.0939	.117	.140	.162	.184	.205	.227	.246	55
56	.0403	.0677	.0923	.115	.138	.159	.180	.201	.223	.242	56
57	.0396	.0665	.0907	.113	.135	.157	.177	.198	.219	.238	57
58	.0389	.0654	.0892	.112	.133	.154	.175	.195	.216	.234	58
59	.0383	.0643	.0877	.110	.131	.152	.172	.191	.212	.230	59
60	.0376	.0633	.0863	.108	.129	.149	.169	.188	209	.226	60
61	.0370	.0623	.0849	.106	.127	.147	.166	.185	.206	223	61
62	.0365	.0613	.0836	.105	.125	.145	.164	.183	.203	.219	62
63	.0359	.0603	.0823	.103	.123	.142	.161	.180	.200	.216	63
64	.0353	.0594	.0810	.101	.121	.140	.159	.177	.197	213	64
65	.0348	.0585	.0798	.0999	.119	.138	.156	.174	.194	.210	65
66	.0343	.0577	.0786	.0984	.117	.136	.154	.172	.191	207	66
67	.0338	.0568	.0775	.0970	.116	.134	.152	.169	.188	.204	67
68	.0333	.0560	.0764	.0956	.114	.132	.150	.167	.185	.201	68
69	.0328	.0552	.0753	.0943	.113	.130	.148	.165	.182	198	69
70	.0324	.0544	.0743	.0930	.111	.128	.146	.162	.179	.195	70

TABLE II—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
71	0319	.0537	.0732	.0917	.109	.127	.144	.160	.177	.193	71
72	.0315	.0530	.0722	0904	.108	.125	.142	.158	.174	.190	72
73	.0310	0522	.0713	.0892	107	.123	.140	156	.172	.188	73
74	.0306	0516	.0703	.0881	.105	.122	.138	.154	.170	.185	74
75	.0302	.0509	.0694	.0869	.104	.120	.136	.152	.167	.183	75
76	.0298	.0502	.0685	.0858	.102	.119	.134	.150	.165	.180	76
77	.0295	.0496	.0676	.0847	.101	.117	.133	.148	.163	.178	77
78	.0291	.0490	.0668	.0836	.0999	.116	.131	.146	161	.176	78
79	0287	0483	.0660	0826	.0987	114	.130	.145	.159	.174	79
80	.0284	.0478	0652	.0816	0974	.113	.128	.143	.157	.172	80
81	0280	.0472	.0644	.0806	.0963	.111	.126	.141	.155	.170	81
82	.0277	.0466	.0636	.0797	0951	.110	.125	.139	.154	.168	82
83	.0274	.0461	.0629	.0787	.0940	.109	.123	.138	.152	.166	83
84	.0270	0455	0621	0778	.0929	.108	.122	.136	.150	.164	84
85	.0267	.0450	.0614	.0769	.0918	.106	.121	.135	.148	.162	85
86	.0264	.0445	.0607	.0760	.0908	.105	.119	.133	.147	.160	86
87	.0261	0440	.0600	.0752	.0898	.104	.118	.132	.145	.158	87
88	.0258	.0435	.0594	.0743	.0888	.103	.117	.130	.143	.157	88
89	.0255	.0430	.0587	0735	0878	.102	.115	.129	.142	.155	89
90	.0253	.0425	.0581	.0727	.0869	.101	.114	.127	.140	.153	90
91	.0250	.0421	.0574	.0719	.0859	.0995	.113	.126	.139	.152	91
92	.0247	.0416	.0568	0712	.0850	.0985	.112	.125	.137	.150	92
93	.0245	.0412	.0562	.0704	.0841	.0974	.110	.123	.136	.148	93
94	.0242	.0408	.0556	.0697	.0832	.0964	.109	.122	.135	.147	94
95	.0239	.0403	.0551	.0690	.0824	.0954	.108	.121	.133	.145	95
96	0237	.0399	0545	.0683	.0815	.0945	.107	.120	.132	.144	96
97	.0235	.0395	.0539	.0676	.0807	.0935	.106	.118	.131	.143	97
98	.0232	.0391	.0534	0669	.0799	.0926	.105	.117	.129	.141	98
99	.0230	0387	.0529	.0662	.0791	.0917	.104	.116	.128	.140	99
100	.0228	.0383	.0524	.0656	.0784	.0908	.103	.115	.127	.138	100
101	.0225	0380	.0518	.0650	.0776	.0899	.102	.114	.125	.137	101
102	.0223	.0376	.0513	.0643	.0768	.0890	.101	.113	.124	.136	102
103	.0221	.0372	0508	.0637	.0761	.0882	.100	.112	.123	.134	103
104	.0219	.0369	.0504	.0631	.0754	.0874	.0991	.111	.122	.133	104
105	.0217	.0365	.0499	.0625	.0747	.0865	.0981	.110	.121	.132	105

TABLE II—*Continued*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
106	.0215	.0362	.0494	.0619	.0740	.0857	.0972	.109	.120	.131	106
107	.0213	.0359	.0490	.0614	.0733	.0850	.0964	.108	.119	.130	107
108	.0211	.0355	.0485	.0608	.0727	.0842	.0955	.107	.118	.128	108
109	.0209	.0352	.0481	.0603	.0720	.0834	.0946	.106	.116	.127	109
110	.0207	.0349	.0477	.0597	.0714	.0827	.0938	.105	.115	.126	110
111	.0205	.0346	.0472	.0592	.0707	.0820	.0930	.104	.114	.125	111
112	.0204	.0343	.0468	.0587	.0701	.0812	.0921	.103	.113	.124	112
113	.0202	.0340	.0464	.0582	.0695	.0805	.0913	.102	.112	.123	113
114	.0200	.0337	.0460	.0577	.0689	.0798	.0906	.101	.111	.122	114
115	.0198	.0334	.0456	.0572	.0683	.0792	.0898	.100	.111	.121	115
116	.0197	.0331	.0452	.0567	.0677	.0785	.0890	.0994	.110	.120	116
117	.0195	.0328	.0449	.0562	.0672	.0778	.0883	.0986	.109	.119	117
118	.0193	.0326	.0445	.0557	.0666	.0772	.0875	.0977	.108	.118	118
119	.0192	.0323	.0441	.0553	.0661	.0765	.0868	.0969	.107	.117	119
120	.0190	.0320	.0437	.0548	.0655	.0759	.0861	.0961	.106	.116	120
121	.0189	.0318	.0434	.0544	.0650	.0753	.0854	.0954	.105	.115	121
122	.0187	.0315	.0430	.0539	.0645	.0747	.0847	.0946	.104	114	122
123	.0185	.0313	.0427	.0535	.0639	.0741	.0841	.0938	.104	.113	123
124	.0184	.0310	.0424	.0531	.0634	.0735	.0834	.0931	.103	.112	124
125	.0183	.0308	.0420	.0527	.0629	.0729	.0827	.0924	.102	.111	125
126	.0181	.0305	.0417	.0523	.0624	.0724	.0821	.0917	.101	.110	126
127	.0180	.0303	.0414	.0519	.0620	.0718	.0815	.0909	100	.110	127
128	.0178	.0301	.0410	.0515	.0615	.0713	.0808	.0902	.0995	.109	128
129	.0177	.0298	.0407	.0511	.0610	.0707	.0802	.0896	.0988	108	129
130	.0176	.0296	.0404	.0507	.0606	.0702	.0796	.0889	.0980	.107	130
131	.0174	.0294	.0401	.0503	.0601	.0696	.0790	.0882	.0973	.106	131
132	.0173	.0291	.0398	.0499	.0596	.0691	.0784	.0876	.0966	.105	132
133	.0172	.0289	.0395	.0495	.0592	.0686	.0778	.0869	.0959	.105	133
134	.0170	.0287	.0392	.0492	.0588	.0681	.0773	.0863	.0952	.104	134
135	.0169	.0285	.0389	.0488	.0583	.0676	.0767	.0857	.0945	.103	135
136	.0168	.0283	.0387	.0485	.0579	.0671	.0762	.0850	.0938	.102	136
137	.0167	.0281	.0384	.0481	.0575	.0666	.0756	.0844	.0931	.102	137
138	.0165	.0279	.0381	.0478	.0571	.0662	.0751	.0838	.0925	.101	138
139	.0164	.0277	.0378	.0474	.0567	.0657	.0745	.0832	.0918	.100	139
140	.0163	.0275	.0376	.0471	.0563	.0652	.0740	.0826	.0912	.0996	140

TABLE II—*Concluded*

n	c										n
	0	1	2	3	4	5	6	7	8	9	
141	.0162	.0273	.0373	.0468	.0559	.0648	.0735	.0821	.0905	.0989	141
142	.0161	.0271	.0370	.0464	.0555	.0643	.0730	.0815	.0899	.0982	142
143	.0160	.0269	.0368	.0461	.0551	.0639	.0725	.0809	.0893	.0975	143
144	0159	.0267	0365	.0458	.0547	.0635	0720	.0804	.0887	.0969	144
145	.0158	0266	.0363	.0455	.0544	.0630	.0715	.0798	.0881	.0962	145
146	.0156	.0264	0360	.0452	0540	.0626	.0710	.0793	0875	.0956	146
147	0155	.0262	.0358	.0449	.0536	.0622	.0705	.0788	.0869	.0949	147
148	.0154	.0260	.0356	.0446	.0533	.0618	.0701	.0783	.0863	.0943	148
149	.0153	0259	0353	.0443	.0529	.0614	.0696	.0777	.0858	.0937	149
150	.0152	.0257	.0351	.0440	.0526	0610	.0692	.0772	.0852	.0931	150

TABLE III

(Based on Poisson approximation to the binomial distribution)

Acceptance Number	Values of $a_1 = np_1$ for which $P(c, a_1) = .95$	Values of $a_2 = np_2$ for which $P(c, a_2) = .10$
0	.05129	2.303
1	.3554	3.890
2	.8177	5.322
3	1 366	6.681
4	1.970	7.994
5	2.613	9.275
6	3.285	10.53
7	3.981	11.77
8	4.695	12.99
9	5.425	14.21
10	6.169	15.41
11	6.924	16.60
12	7.690	17.78
13	8.464	18.96
14	9.246	20.13
15	10.04	21.29

REFERENCES

[1] CATHERINE M THOMPSON, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32 (1944), pp. 151–181.

[2] MAXINE MERRINGTON AND CATHERINE M. THOMPSON, "Tables of percentage points of the inverted beta (F) distribution," *Biometrika*, Vol. 33 (1945), pp 77–88

[3] PAUL PEACH AND S. B. LITTAUER, "A note on sampling inspection," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 81–84.

ON THE RANGE-MIDRANGE TEST AND SOME TESTS WITH BOUNDED SIGNIFICANCE LEVELS[1]

By John E. Walsh

The RAND Corporation

1. Summary. This paper is divided into two parts. The significance tests investigated in Part I concern the population mean and are based on the quantity

[(sample midrange)-(hypothetical mean)]/(sample range).

The case in which the observations are a sample from a normal population is considered in detail. The tests investigated are summarized in Table 1. These tests are found to be very efficient for small samples (see Table 4, power efficiency is defined in section 3). An investigation of several extremely non-normal populations using the values of $D\alpha$ obtained for normality indicates that the significance level of the range-midrange test is not very sensitive to the requirement of normality for small samples (see Table 6). Also the tests of Table 1 can be applied without computation through the use of an easily constructed graph (see section 4). These properties suggest that the range-midrange test is preferable to the Student t-test and the analogue of the Student t-test using the sample range (see [1] and [2]) whenever the sample size is sufficiently small.

Use of the range-midrange test for the case of normality was proposed by E. S. Pearson in [3], where properties of the test were experimentally investigated for the normal and certain non-normal populations.

In Part II several significance tests for the mean are developed which have a specified significance level for the case of a sample from a normal population but whose significance level is bounded near the specified value under very general conditions, one of which is that the observations are from continuous symmetrical populations. Some of these tests are range-midrange tests. Table 2 contains a summary of the tests and their properties (x_i = ith largest observation, $i = 1, \cdots, n$; conditions (D) are given in section 7).

PART I. THE RANGE-MIDRANGE TEST

2. Introduction. In 1929 E. S. Pearson proposed using the range-midrange test for the case of a sample from a normal population (see [3]) and experimentally investigated some of its properties for sample sizes of 5 and 10 and significance levels of 2% and 10% (symmetrical tests). Using the constants (corresponding to the D_α in this paper) determined for the case of normality,

[1] This paper was presented to a joint meeting of the Institute of Mathematical Statistics and the American Mathematical Society at New Haven, Conn. in September, 1947. The results presented in this paper were obtained in the course of research conducted under the sponsorship of the Office of Naval Research. This research was performed while the author was at Princeton University.

significance level and power function properties of these four tests were experimentally investigated for several non-normal populations. The results of this empirical investigation indicated that the range-midrange test is very efficient for normality and not very sensitive to the assumption of normality if the sample size is sufficiently small.

This paper presents an analytical investigation of properties of the range-midrange test for $n = 2, 3, \cdots, 10$ and a wide range of significance levels. The results of this investigation confirm the contention that the range-midrange test is very efficient for normality and small samples; also an analytical investigation of how the significance level changes for the case of certain extremely non-normal populations furnishes results which agree with the contention that the range-midrange test is not very sensitive to the requirement of normality for sufficiently small samples.

In most cases the results presented in this paper are not directly comparable with those obtained by Pearson. It was possible, however, to obtain values of $D\alpha$, ($\alpha = 5\%, 1\%$; $n = 5, 10$), from the results presented in [3]; these values were found to be in close agreement with the corresponding values of Table 5.

3. Efficiency of range-midrange. The purpose of this section is to use the relations derived in section 6 to determine the power efficiencies of tests A, B and C (see Table 1) for $\alpha = 1\%, 5\%$ and $n = 2, \cdots, 10$. To do this the method of defining power efficiency given in [4] and [5] will be used. As shown in [5], it is sufficient to consider only test A; for any fixed n and α, tests A, B and C all have the same power efficiency (note that the significance level of test C is 2α).

For a normal population (unknown variance) the most powerful test of the one-sided alternative $\mu < \mu_0$ is the appropriate Student t-test. The procedure used in determining the power efficiency of test A consists in first computing the power function of test A for the given values of n and α; then the sample size of the corresponding Student t-test at this significance level is varied until the power function of the t-test is approximately equal to that of test A. The size sample (not necessarily integral) thus obtained for the t-test divided by n is called the power efficiency of test A for the given values of n and α. Intuitively the power efficiency of a test measures the percentage of the total available information per observation which is being utilized by that test.

Table 3 contains values of the power function for test A. These values were computed from equation (3) of section 6 by approximate integration.

The corresponding values of the power function for the Student t-test were found by using the normal approximation given in [6]. This approximation was used for fractional degrees of freedom. The sample sizes considered as well as the resulting power function values are listed in Table 3. A comparison of the power function values for the two types of tests furnishes the approximate power efficiencies listed in Table 3.

For $n = 2$, test A is itself a Student t-test. The power efficiency is therefore 100% for that sample size. This combined with Table 3 furnishes power

efficiencies at the 1% level for $n = 2, 6, 8, 10$ and at the 5% level for $n = 2, 6, 10$. The approximate power efficiencies given in Table 4 for other values of n were obtained from these values by graphical interpolation.

Table 4 shows that the power efficiency for $\alpha = 1\%$ is very good for $n \leq 8$, while for $\alpha = 5\%$ the efficiency is good for $n \leq 6$.

TABLE 1

Summary of range-midrange tests

Definitions	Tests		Significance Level
	Accept	If	
Test based on sample of size n, $(2 \leq n \leq 10)$, from an arbitrary normal population.	(A)		
. . . . .	$\mu < \mu_0$	$D < -D_\alpha$	α
x_1 = smallest sample value.			
x_n = greatest sample value.			
. . . . .			
μ = the mean of the normal population.	(B)		
μ_0 = given hypothetical mean value to be tested.			
. . . . .	$\mu > \mu_0$	$D > D_\alpha$	α
$D = \dfrac{\text{(sample midrange)-(hypothetical mean)}}{\text{(sample range)}}$			
$= [(x_n + x_1)/2 - \mu_0]/(x_n - x_1)$.			
. . . . .	(C)		
D_α = constant depending on n and α. Values of α versus D_α for $2 \leq n \leq 10$ and $\alpha = 5\%, 2.5\%, 1\%, 0.5\%$ are given in Table 5.	$\mu \neq \mu_0$	$\lvert D \rvert > D_\alpha$	2α

4. Construction of graph. In most problems to which a test of the type developed in this paper would be applied, the values of the sample can be considered to have practical lower and upper limits, say a and b. For example, in many situations zero is a lower limit for the sample values. From a practical viewpoint these limits on the sample values do not contradict the assumption that the population is normal, since the area under that part of the normal distribution which lies outside the interval (a, b) can be considered negligible. Thus, since $\Pr(u/v^2 \leqslant w) = \Pr(u \leqslant v^2 w)$, test A can be restated in the form

Accept $\mu < \mu_0$ if the sample point (x_1, x_n) falls in the region (A) of the x_1, x_n

TABLE 2

Some one-sided and symmetrical tests with bounded significance levels

n	Tests		Significance Level for Normality		Significance Level Bounds for Conditions (D)				Approx Efficiency for normality
	SYMMETRICAL: Accept $\mu \neq \mu_0$ if either				1-sided Tests		Symmet Tests		
	ONE-SIDED: Accept $\mu < \mu_0$ if	*ONE-SIDED:* Accept $\mu > \mu_0$ if	One-sided	Symmetrical	Upper	Lower	Upper	Lower	
			%	%	%	%	%	%	%
4	$1.055x_4 - .055x_1 < \mu_0$	$1.055x_1 - .055x_4 > \mu_0$	5	10	6.2		12.5		96
5	$.63x_5 + .37x_4 < \mu_0$	$.63x_1 + .37x_2 > \mu_0$	5	10	6.2	3.1	12.5	6.2	99
	$1.02x_5 - .02x_1 < \mu_0$	$1.02x_1 - .02x_5 > \mu_0$	2.5	5	3.1		6.2		97
6	$.63x_6 + .37x_5 < \mu_0$	$.63x_1 + .37x_2 > \mu_0$	2.5	5	3.1	1.6	6.2	3.1	98.5
	$1.06x_6 - .06x_1 < \mu_0$	$1.06x_1 - .06x_6 > \mu_0$	1	2	1.6		3.1		98
7	$.785x_7 + .215x_6 < \mu_0$	$.785x_1 + .215x_2 > \mu_0$	1	2	1.6	0.8	3.1	1.6	97
	$1.05x_7 - .05x_1 < \mu_0$	$1.05x_1 - .05x_7 > \mu_0$	0.5	1	0.8		1.6		96
8	$\max[x_7, (.5x_8 + .28x_6 + .22x_7)] < \mu_0$	$\min[x_2, (.5x_1 + .28x_3 + .22x_2)] > \mu_0$	Approx. 1	Approx. 2	1.2	0.8	2.3	1.6	98
	$.785x_6 + .215x_7 < \mu_0$	$.785x_1 + .215x_2 > \mu_0$	0.5	1	0.8	0.4	1.6	0.8	97
9	$\max[x_8, (.5x_9 + .28x_7 + .22x_8)] < \mu_0$	$\min[x_2, (.5x_1 + .28x_3 + .22x_2)] > \mu_0$	Approx. 0.5	Approx. 1	0.6	0.4	1.2	0.8	98.5

TABLE 3

Power function values for test A

Type Test	Sample Size	Approx. Efficiency	Significance Level	Approximate Values of Power Function				
				$\delta = \frac{1}{2}$	$\delta = 1$	$\delta = 1\frac{1}{2}$	$\delta = 2$	$\delta = 2\frac{1}{2}$
		%						
t	5.4		.05	.244	.607	.886	.969	
A	6	90	.05	.259	.599	.868	.967	
t	7.5		.05	.333	.783	.971		
A	10	75	.05	.351	.779	.962		
t	5.88		.01	.071	.248	.551	.820	.957
A	6	98	.01	.077	.271	.568	.809	.935
t	7.2		.01	.091	.371	.749	.949	
A	8	90	.01	.104	.389	.728	.923	
t	8		.01	.108	.453	.832	.976	
A	10	80	.01	.124	.462	.814	.963	

TABLE 4

Power efficiencies of tests A, B and C for $\alpha = 5\%, 1\%$ and $2 \leq n \leq 10$

α	n								
	2	3	4	5	6	7	8	9	10
.01	100%	99.7%	99.4%	99%	98%	95%	90%	85%	80%
.05	100%	98.5%	96%	93.5%	90%	86.5%	82.5%	78.5%	75%

TABLE 5

Approximate values of D_α for $\alpha = 5\%, 2.5\%, 1\%, 0.5\%$ and $2 \leq n \leq 10$

α	n								
	2	3	4	5	6	7	8	9	10
0.5%	31.83	3.02*	1.37*	.85*	.66	.55*	$.47_5$	$.42_5$	.39*
1%	15.91	2.11*	1.04*	.71	.56*	$.47_5$	.42*	.38	.35*
2.5%	6.35	1.30	.74	.52	.43	$.37_5$	.33	.30	$.27_5$
5%	3.16	.90*	$.55_5$*	$.42_5$	.35*	.30	$.26_5$	.24	$.22_5$*

* These values of D_α were verified directly by substitution and integration. The remaining values of D_α for $3 \leqq n \leqq 10$ were obtained from these and other values of D_α, ($\alpha \pm .005, .01, .025, .05$), by graphical interpolation.

plane defined by

$$(1/2 + D_\alpha)x_n + (1/2 - D_\alpha)x_1 < \mu_0\,, \qquad x_n \geq x_1\,, \qquad a \leq x_1\,, x_n \leq b.$$

TABLE 6

Effect of non-normality on the significance level of the range-midrange test

n	Probability Density Function	Significance Level											
		Test A				Test B				Test C			
	Normal	.05	.025	.01	.005	.05	025	.01	.005	.10	.05	.02	.01
3	1 if $0 \leq x \leq 1$	.064	.039	.018	.010	064	.039	.018	.010	.128	.078	.036	020
4	0 otherwise	.053	.033	.017	.0096	.053	.033	.017	.0096	.106	.066	.034	.0192
5	Mean = $\frac{1}{2}$	.043	029	015	.0094	.043	029	.015	.0094	.086	.058	030	.0188
3	$\frac{1}{2}e^{\lvert x\rvert}$, $-\infty < x < \infty$	.036	.017	.0063	.0031	036	.017	.0063	.0031	.072	.034	.0126	.0062
4	Mean = 0	.043	.016	.0055	.0024	.043	016	.0055	.0024	.086	.032	.0101	.0048
5		.095	.026	.0059	.0027	.095	.026	.0059	.0027	.190	052	.0118	0054
3	$\frac{3}{2}x^2$ if $-1 \geq x \geq 1$	119	.104	.073	.050	.119	.104	.073	.050	.238	.208	.146	.100
4	0 otherwise	.062	.061	055	.045	.062	061	.055	045	.124	.122	.110	.090
5	Mean = 0	.031	.031	.031	.029	.031	.031	.031	.029	.062	.062	.062	.058
3	e^{-x} if $0 \leq x < \infty$	.014	.0067	.0025	.0012	.158	.108	.059	.035	.172	.115	.062	.036
4	0 otherwise	013	0048	.0016	0007	.144	.104	.065	042	157	.109	.067	.043
5	Mean = 1	.017	.0055	.0013	0006	.122	.096	.061	.045	.139	.102	.062	.046
3	$2x$ if $0 \leq x \leq 1$	035	.019	.0075	.0038	.096	061	.030	.017	.131	.080	.038	.021
4	0 otherwise	031	.016	0065	.0031	.083	055	.031	.018	.114	.071	.038	021
5	Mean = $\frac{2}{3}$	.028	015	.0057	.0031	.068	.050	.028	.019	.096	.065	.032	.020
3	$3x^2$ if $0 \leq x \leq 1$	.027	014	.0053	.0026	.112	072	037	.021	.139	.086	.042	.024
4	0 otherwise	024	.011	.0043	.0019	.099	.067	.039	.024	.123	.078	.043	.026
5	Mean = $\frac{3}{4}$	.023	.012	.0037	.0019	082	.061	.036	.025	.105	.073	.040	.027

Likewise test B can be restated as

Accept $\mu > \mu_0$ if (x_1, x_n) falls in the region (B) defined by

$$(1/2 - D_\alpha)x_n + (1/2 + D_\alpha)x_1 > \mu_0\,, \qquad x_n \geq x_1\,, \qquad a \leq x_1\,, x_n \leq b.$$

Test C now becomes

Accept $\mu \neq \mu_0$ *if* (x_1, x_n) *falls in either of the regions* (A) *or* (B).

Figure 1 (i) contains a schematic diagram of the regions (A) and (B). Test A can be applied by constructing a graph of the region (A) and giving the instructions to accept $\mu < \mu_0$ if (x_1, x_n) falls in (A). Similarly for test B and region (B). Test C is applied by constructing a graph of both (A) and (B) and accepting $\mu \neq \mu_0$ if (x_1, x_n) falls in either (A) or (B).

Frequently it is desirable to simultaneously consider more than one significance level. This can be accomplished in the manner indicated by Figure 1(ii).

5. Effect of non-normality on significance level. It has been shown that the range-midrange test compares very favorably with the Student t-test for suffi-

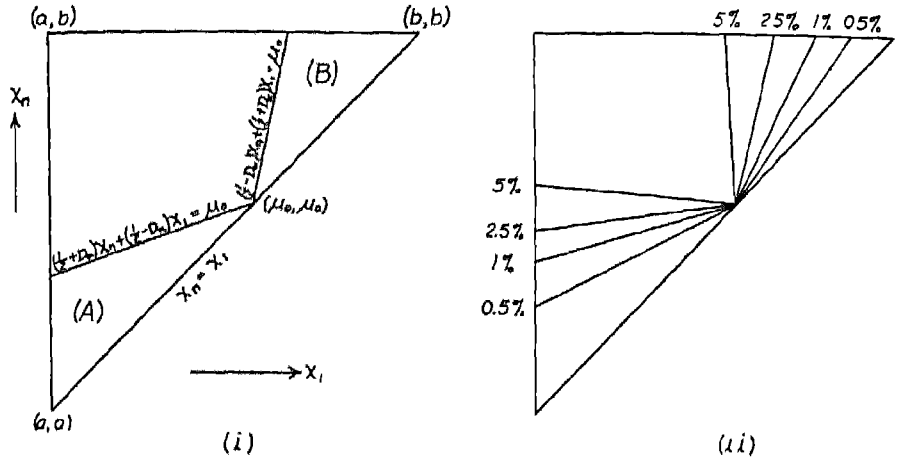

FIG. 1. SCHEMATIC DIAGRAMS OF REGIONS USED IN CONSTRUCTION OF GRAPHS

ciently small samples and normality. In practice, however, it may happen that normality is assumed for cases in which the population is not even approximately normal. Although this represents an error in judgment on the part of the person applying the test, such situations will undoubtedly occur if the range-midrange test is used very frequently. The purpose of this section is to investigate the effect of non-normality on the significance level of the range-midrange test when the values of D_α based on normality are used. The corresponding effect of these non-normal populations on the significance level of the t-test was not considered because of computational difficulties; however the effect of some other non-normal populations on the significance level of the t-test was experimentally investigated by Pearson in [3]. The results of this empirical investigation and of later investigations shows that the significance level of the t-test is not very sensitive to the requirement of normality for small samples.

Six populations were chosen for investigation. Three of these populations are

symmetrical while the remaining three are strongly asymmetrical. These particular populations were considered because their probability density functions have a wide variety of different shapes; also because the significance level of the range-midrange test can be computed in closed form for these populations.

The populations investigated are defined by their probability density functions. Table 6 contains a list of the probability density functions considered along with the resulting significance levels for the range-midrange test. The cases investigated are $n = 3, 4, 5$ and $\alpha = 5\%, 2.5\%, 1\%, 0.5\%$. Larger values of n were not used because of computational difficulties. The situation of $n = 2$ was not considered because the t-test and the range-midrange test are identical for this case. The significance levels of Table 6 were computed by making direct application of (1) and (2) of section 6.

6. Significance level and power function derivations. The purpose of this section is to present derivations of the significance level and power function expressions which were used in the preceding sections. First a general probability expression will be evaluated. Direct applications of the results obtained for this expression yield the required significance level and power function relations.

Let x_1 and x_n be the smallest and largest values, respectively, of a sample of size n drawn from a population with probability density function $f(x)$ The non-zero probability range of this population is $\gamma \leq x \leq \beta$. Also let three constants c_1, c_n, c_0, $(c_1 + c_n = 1)$, be given and consider the value of

$$\Pr(c_1 x_1 + c_n x_n < c_0); \quad \text{where} \quad M(z) = \int_{-\infty}^{z} f(y)\, dy.$$

Using direct methods it is found that the value of this expression is given by

$$(1)\quad \begin{cases} [M(c_0)]^n & \text{if } c_1 = 0. \\ 0 & \text{if } 0 < c_1 < 1,\ c_0 \leq \gamma \\ M\left(\dfrac{c_0 - c_1\gamma}{c_n}\right)^n - n\displaystyle\int_{c_0}^{(c_0 - c_1\gamma)/c_n} \left[M(V) - M\left(\dfrac{c_0 - c_n V}{c_1}\right)\right]^{n-1} f(V)\, dV & \\ \qquad \text{if } 0 < c_1 < 1, \quad c_0 > \gamma. & \\ 1 - [1 - M(c_0)]^n & \text{if } c_1 = 1 \\ 0 & \text{if } c_1 > 1,\ c_0 \leq \min[\gamma, c_1\gamma + c_n\beta]. \\ 1 - n\displaystyle\int_{(c_0 - c_n\gamma)/c_n}^{\beta} \left[M(V) - M\left(\dfrac{c_0 - c_n V}{c_1}\right)\right]^{n-1} f(V)\, dV & \\ \qquad - M\left(\dfrac{c_0 - c_n\gamma}{c_n}\right)^n \quad \text{if } c_1 > 1, \quad c_1\gamma + c_n\beta < c_0 \leq \gamma. & \\ 1 - n\displaystyle\int_{c_0}^{\beta} \left[M(V) - M\left(\dfrac{c_0 - c_n V}{c_1}\right)\right]^{n-1} f(V)\, dV \quad \text{if } c_1 > 1, \quad c_0 > \gamma. & \end{cases}$$

The value of $\Pr(c_1x_1 + c_nx_n < c_0)$ for $c_1 < 0$ can be obtained from the above results for $c_1 > 1$. It is easily shown that

$$\Pr(c_1x_1 + c_nx_n < c_0) = 1 - \Pr(c_1'y_1 - c_n'y_n < c_0'), \tag{2}$$

where

$$c_1' = c_n, \qquad c_n' = c_1, \qquad c_0' = -c_0,$$

and y_1, y_n are the smallest and largest values, respectively, of a sample of size n drawn from a population with probability function $g(y) = f(-y)$. Thus if $c_1 < 0$, $c_1' = c_n > 1$ and obvious modifications of the results for $c_1 > 1$ will furnish the value of $\Pr(c_1'y_1 + c_n'y_n < c_0')$.

The above general results were used in section 5 to investigate the effect of non-normality on the significance level of the range-midrange test

Now consider the case in which the n sample values are drawn from a normal population with mean μ and variance σ^2. Then, for test A,

$$\begin{aligned}\text{Power Function} &= \Pr\{(1/2 - D_\alpha)x_1 + (1/2 + D_\alpha)x_n < \mu_0\} \\ &= \Pr\{(1/2 - D_\alpha)z_1 + (1/2 + D_\alpha)z_n < \delta\},\end{aligned}$$

where

$$z_1 = (x_1 - \mu)/\sigma, \quad z_n = (x_n - \mu)/\sigma, \quad \delta = (\mu_0 - \mu)/\sigma.$$

Using the above results with

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad M(z) = N(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx,$$

it is found that the power function for test A is

$$\begin{cases} 1 - n\displaystyle\int_{\delta}^{\infty} \left[N(V) - N\left\{\frac{\delta - (1/2 + D_\alpha)V}{1/2 - D_\alpha}\right\}\right]^{n-1} f(V)\, dV & \text{if } D_\alpha < 1/2; \\ [N(\delta)]^n \quad \text{if} \quad D_\alpha = 1/2; \\ n\displaystyle\int_{-\infty}^{\delta} \left[N(V) - N\left\{\frac{\delta - (1/2 + D_\alpha)V}{1/2 - D_\alpha}\right\}\right]^{n-1} f(V)\, dV, & \text{if } D_\alpha > 1/2. \end{cases} \tag{3}$$

The value of D_α (for given n) corresponding to a specified significance level α for test A is obtained by solving the equation

$$\alpha = P_A(0), \tag{4}$$

where $P_A(\delta)$ is the power function for test A. From symmetry and the fact that test C is a combination of tests A and B, test B has significance level α and test C significance level 2α for this value of D_α.

For $n = 2$, test A becomes a Student t-test with one degree of freedom if D_α is replaced by $t_\alpha/2$. The relation $D_\alpha = t_\alpha/2$ gives an easily applied method of computing D_α for this case.

Approximate values of D_α for $\alpha = 5\%$, 2.5%, 1%, 0.5% are contained in

Table 5 for $2 \leq n \leq 10$. For $3 \leq n \leq 10$, these values were obtained from (3) and (4) by approximate integration and interpolation. For $n = 2$, the relation between D_α and t_α was used.

PART II. SOME TESTS WITH BOUNDED SIGNIFICANCE LEVELS

7. Introduction. In this part some significance tests (for the mean) are derived which are based on the assumption of a sample from a normal population. These tests have the property that the significance level is bounded near the value for normality under very general conditions. These conditions are

(D) (a) The observations used for a test are independent.
(b) Each observation comes from a continuous symmetrical population with mean μ.

It is to be emphasized that no two observations are necessarily drawn from the same population

The bounded significance level tests developed are summarized in Table 2. These tests can be used to supplement the tests presented in [5] for $n \leq 9$, where the tests of [5] do not furnish a very wide variety of suitable significance levels.

8. Outline of derivations. Let us consider the range-midrange test for the more general situation in which the set of independent observations used are from arbitrary but fixed populations satisfying conditions (D). Let D_α be redefined so that the resulting test A has significance level α. Then it is easily seen that D_α is a monotone decreasing function α. Thus the significance level of the modified test A will always be less than or equal to $(1/2)^n$ if $D_\alpha > 1/2$. The significance level bounds for the tests $n = 4$, $\alpha = 5\%$; $n = 5$, $\alpha = 2.5\%$; $n = 6$, $\alpha = 1\%$, $n = 7$, $\alpha = 0.5\%$ of Table 2 were obtained from this relation and obvious significance level relations among tests A, B and C.

The significance levels (for normality) for the tests $n = 5$, $\alpha = 5\%$; $n = 6$, $\alpha = 2.5\%$; $n = 7$, $\alpha = 1\%$; $n = 8$, $\alpha = 0.5\%$ were obtained by approximate integration of the expression derived for $\Pr[(1/2 + c)x_n + (1/2 - c)x_{n-1} < \mu]$, $(0 < c < 1/2)$, for several values of c and then graphical interpolation (here α is the one-sided test significance level). The significance level bounds were determined from

$$(1/2)^n = \Pr(x_n < \mu) \leq \Pr[(1/2 + c)x_n + (1/2 - c)x_{n-1} < \mu]$$
$$\leq \Pr[(1/2)(x_n + x_{n-1}) < \mu] = (1/2)^{n-1}$$

The significance levels for the tests $n = 8$, $\alpha = 1\%$; $n = 9$, $\alpha = 0.5\%$ were obtained by considering the relations

$$\Pr\{\max [x_{n-1}, (x_n + x_{n-1})/2] < \mu\} = (1 + i)(1/2)^n, \qquad (i = 0, 1, 2, 3),$$

and applying linear interpolation to find a value c, $(0 < c < 1/2)$, such that

$\Pr\{\max [x_{n-1}, 0.5x_n + cx_{n-2} + (1/2 - c)x_{n-1}] < \mu\}$ has the desired value. The significance level bounds were found from

$$\Pr\{(1/2)(x_n + x_{n-1}) < \mu\} \leq \Pr\{\max[x_{n-1}, 0.5x_n + cx_{n-2} + (\tfrac{1}{2} - c)x_{n-1}] < \mu\}$$
$$\leq \Pr\{\max[x_{n-1}, (1/2)(x_n + x_{n-2})] < \mu\}.$$

The derivation of the power efficiencies listed in Table 2 will not be considered here. Detailed derivations can be found in [7].

REFERENCES

[1] J. F. Daly, "On the use of the sample range in an analogue of Students' t-test," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71–74.

[2] E. Lord, "The use of range in place of standard deviation in t-test," *Biometrika*, Vol. 34 (1947), pp. 41–67.

[3] E. S. Pearson, "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations," *Biometrika*, Vol. 21 (1929), pp. 280–286.

[4] John E. Walsh, "On the power function of the sign test for slippage of means," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 358–362.

[5] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 64–81.
610–611. Submitted for publication in *Annals of Math. Stat.*

[6] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), p. 376.

[7] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *unpublished thesis, Princeton University Library*, Princeton, N. J.

ASYMPTOTIC STUDENTIZATION IN TESTING OF HYPOTHESES

By Herman Chernoff[1]

Cowles Commission for Research in Economics

1. Summary. A method suggested by Wald for finding critical regions of almost constant size and various modifications are considered. Under reasonable conditions the sth step of this method gives a critical region of size $\alpha + R_s(\theta)$ where θ is the unknown value of the nuisance parameter, $R_s(\theta) = O(N^{-s/2})$ and N is the sample size. The first step of this method gives the region which is obtained by assuming that an estimate $\hat{\theta}$ of the nuisance parameter is actually equal to θ.

2. Introduction. The problem of nuisance parameters often arises in the testing of hypotheses in the following form: It is desired to construct a test of a hypothesis H so that the probability of rejecting H if it is true is equal to α. However the probability distribution of the data is not uniquely determined by H. Indeed, if the hypothesis is true then the observations have a distribution depending on a nuisance parameter θ whose value is not known. Generally a critical region will have a size which depends on the value of θ. Neyman has done considerable work on the problem of finding similar regions, i.e., regions whose size is independent of θ.

Wald has suggested the following method of finding critical regions whose size is almost independent of θ. Suppose that t is a statistic such that if θ were known then the critical region $t \leq c_1(\theta)$ would be a good critical region for testing the hypothesis H. Suppose also that $\hat{\theta}$ is an estimate of θ and that $g(t, \hat{\theta} \mid \theta)$ represents the joint distribution of t, $\hat{\theta}$ under H when θ is the value of the nuisance parameter. Then consider the regions

$$t \leq c_1(\hat{\theta}) \quad \text{where} \quad \Pr\{t \leq c_1(\theta)\} = \alpha \qquad \text{independent of } \theta;$$

$$t \leq c_1(\hat{\theta}) + c_2(\hat{\theta}) \quad " \quad \Pr\{t - c_1(\hat{\theta}) \leq c_2(\theta)\} = \alpha \text{ independent of } \theta;$$

$$t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta}) \quad " \quad \Pr\{t - c_1(\hat{\theta}) \cdots - c_{s-1}(\hat{\theta}) \leq c_s(\theta)\} = \alpha$$

independent of θ.

Under the assumption that $\hat{\theta}$ is close to θ it is reasonable to expect that $\Pr\{t \leq c_1(\hat{\theta})\}$ would be close to α. It might also be expected that $\Pr\{t \leq c_1(\hat{\theta}) + c_2(\hat{\theta})\}$ would be even closer to α.

This method has been shown to have good properties when considered from the asymptotic point of view. Suppose that t, $\hat{\theta}$ are two sequences of statistics

[1] This paper is based on a dissertation written under the supervision of Professor Abraham Wald and submitted as partial fulfilment of the requirements for Ph.D. in the Graduate Division of Applied Mathematics of Brown University.

(depending on N, the size of the sample or an analogous variable) with distribution represented by $g(t, \hat{\theta} \mid \theta)$ where N is understood to be present. Then it has been shown that under reasonable conditions, with modifications for the sake of calculation,

$$| \Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha | = O(N^{-s/2})$$

The statement of the theorem presenting this result will be given in section 4. It has also been shown that if roughly speaking $\hat{\theta}$ is distributed almost symmetrically about θ, the above result may be obtained in half the steps, i.e.,

$$| Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha | = O(N^{-s}).$$

It is true that under relatively weak conditions and for fixed N it is possible for any $\epsilon > 0$ to obtain a function $h(\hat{\theta})$ such that $| \Pr\{t \leq h(\hat{\theta})\} - \alpha | < \epsilon$. However such a critical region can have very poor properties from the point of view of the alternative hypotheses especially if $h(\hat{\theta})$ is a very wildly oscillating function. On the other hand this objection does not apply to Wald's method for large N because

$$\begin{aligned} | c_1^{(r)}(\theta) | &\leq M & r &= 0, 1, \cdots, s; \\ | c_2^{(r)}(\theta) | &\leq MN^{-1/2} & r &= 0, 1, \cdots, s-1; \\ | c_s^{(r)}(\theta) | &\leq MN^{-(s-1)/2} & r &= 0, 1, \end{aligned}$$

and hence $c_1(\hat{\theta}) + \cdots + c_2(\hat{\theta})$ is almost constant over "that small range in which $\hat{\theta}$ will probably fall."

In the above it has been implied that θ is a one dimensional variable. However the results are easily extended to the case where θ is a k-dimensional variable.

The direct application of the method is often quite difficult because of the calculations involved. Modifications can be applied which simplify the calculations. Such modification usually consist of changing the $c_r(\theta)$ by a small amount provided the remainder is simple and "well behaved." A case where considerable simplifications can be made is that where $g_1(t \mid \hat{\theta}, \theta)$, the conditional distribution of t, can be expanded in a Taylor Expansion,

$$g_1(t \mid \hat{\theta},\theta) = g_1(c_1(\theta) \mid \theta, \theta) + (t - c_1(\theta)) \frac{\partial g_1}{\partial t} + (\hat{\theta} - \theta) \frac{\partial g_1}{\partial \theta}$$

$$+ \cdots + \frac{1}{s!} \sum_{j=0}^{s} (t - c_1(\theta))^j (\hat{\theta} - \theta)^{s-j} \frac{\partial^s}{\partial t^j \partial \hat{\theta}^{s-j}} g_1(t' \mid \hat{\theta}', \theta),$$

where the partial derivatives "behave." This case will be described in detail in section 3, and an example previously treated by Welch (see [1]) will be discussed in section 4.

Another case where simplifications often arise is the asymptotic case, that is the case where $g(t, \hat{\theta} \mid \theta)$ has an asymptotic expansion. The asymptotic case

may also be regarded as an extension of the following partition principle which is very useful. If $g(t, \hat{\theta} \mid \theta) = g_0(t, \hat{\theta} \mid \theta) + h(t, \hat{\theta} \mid \theta)$ and $\iint \mid h \mid dt\, d\hat{\theta} \leq MN^{-s/2}$ and if $\varphi(\hat{\theta})$ is such that

$$\left| \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{\varphi(\theta)} dt\, g_0(t, \hat{\theta} \mid \theta) - \alpha \right| \leq MN^{-s/2},$$

then $\mid \Pr\{t \leq \varphi(\hat{\theta})\} - \alpha \mid \leq MN^{-s/2}$. Thus our theorems apply to $g(t, \hat{\theta} \mid \theta)$ if $g = g_0 + h$ where g_0 has sufficient differentiability properties.

3. The Taylor expansion treatment. Let $g(t, \hat{\theta} \mid \theta) = g_1(t \mid \hat{\theta}, \theta) g_2(\hat{\theta} \mid \theta)$ where g_1 is the conditional density of t given $\hat{\theta}$ and $g_2(\hat{\theta} \mid \theta)$ is the marginal density of $\hat{\theta}$. $g_3(t \mid \theta) = \int_{-\infty}^{\infty} d\hat{\theta} g(t, \hat{\theta} \mid \theta)$ is the marginal density of t. In what follows we shall use M as a generic bound. Thus the statement $f(t, \theta) < M(\theta_1, \theta_2)$, $\theta_1 \leq \theta \leq \theta_2$, means that there is a constant M depending on (θ_1, θ_2) and independent of t, θ, N so that $f(t, \theta) < M(\theta_1, \theta_2)$ $\theta_1 \leq \theta \leq \theta_2$.

First we obtain $c_1(\theta)$ so that $\Pr\{t \leq c_1(\theta)\} = \alpha$.
Then we have

THEOREM 1. *If for every finite interval* (θ_1, θ_2),

(i) $\left| \frac{\partial^p g_3}{\partial \theta^p} (t \mid \theta + \Delta) \right| < G_1(t, \theta) < G_2(t), \quad \mid \Delta \mid \leq \Delta'(\theta_1, \theta_2, N), \quad p = 0, 1, \cdots s,$

$$\theta_1 \leq \theta,\ \theta + \Delta \leq \theta_2,$$

where $\int_{-\infty}^{\infty} G_2(t)\, dt < M(\theta_1, \theta_2)$, *$G_1$ and G_2 may depend on N, θ_1, and θ_2*

(ii) $\frac{\partial^{p+q} g_3(t \mid \theta)}{\partial \theta^p \partial t^q}$ *is continuous in t, θ and*

bounded in absolute value by $M(C_1, C_2, \theta_1, \theta_2)$ for $p + q < s$, $\theta_1 \leq \theta \leq \theta_2$, $C_1 \leq t \leq C_2$;

(iii) $0 < \frac{1}{M(C_1, C_2, \theta_1, \theta_2)} < g_3(t \mid \theta)$ *for* $\theta_1 \leq \theta \leq \theta_2$, $C_1 \leq t \leq C_2$;

(iv) $0 < \alpha < 1$,

then $\Pr\{t \leq c_1(\theta)\} = \alpha$ *defines $c_1(\theta)$ uniquely and so that* $\mid c_1^{(p)}(\theta) \mid \leq M(\theta_1, \theta_2)$ *for* $p = 0, 1, \cdots, s$ $\theta_1 \leq \theta \leq \theta_2$.

PROOF. Since $g_3(t \mid \theta)$ is positive, $c_1(\theta)$ is uniquely defined by condition (i). From this and conditions (i) and (ii) it follows that $c_1'(\theta)$ exists and is given by

$$\int_{-\infty}^{c_1(\theta)} dt \frac{\partial g_3}{\partial \theta} (t \mid \theta) + c_1'(\theta) g_3(c_1(\theta) \mid \theta). \tag{1}$$

We may continue in this fashion differentiating formally $p \leq s$ times to get

$$(2)\qquad \int_{-\infty}^{c_1(\theta)} dt\, \frac{\partial^p g_3(t \mid \theta)}{\partial \theta^p} + \sum [c_1^{(j_1)}(\theta)]^{i_1}[c_1^{(j_2)}(\theta)]^{i_2} \cdots [c_1^{(j_k)}(\theta)]^{i_k} \frac{\partial^{i+j} g_3(c_1(\theta) \mid \theta)}{\partial \theta^i \partial t^j}$$
$$+ c_1^{(p)}(\theta) g_3(c_1(\theta) \mid \theta) = 0, \qquad j_1, j_2, \cdots, j_k, i_1, \cdots, i_k, i + j < p.$$

From the continuity and positiveness it follows that $c^{(p)}(\theta)$ is continuous. Since $\int_{-\infty}^{\infty} G_2(t)\, dt < M(\theta_1, \theta_2)$ it follows that there is a constant $M(\theta_1, \theta_2)$ so that

$$\int_{-\infty}^{-M(\theta_1,\theta_2)} G_2(t)\, dt < \alpha, \qquad \int_{M(\theta_1,\theta_2)}^{\infty} G_2(t)\, dt < 1 - \alpha.$$

Thus

$$| c_1(\theta) | \leq M(\theta_1, \theta_2).$$

From (1) and condition (i) it follows easily that $| c_1'(\theta) | \leq M(\theta_1, \theta_2)$. Similarly we obtain $| c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2)$ for $\theta_1 \leq \theta \leq \theta_2$.

While the conditions (i) to (iv) suffice to insure the results of the theorem they are not necessary. It is often possible to obtain these properties of $c_1(\theta)$ in particular examples where $g_3(t, \theta)$ does vanish at points so long as $g_3(c_1(\theta), \theta)$ behaves well.

DEFINITION 1. *$\varphi_m(\hat{\theta})$ is an admissible function of order $m (m \leq s$, s fixed in advance) if $\varphi_m(\hat{\theta}) = c_1(\hat{\theta}) + \cdots + c_m(\hat{\theta})$ where $\Pr\{t \leq c_1(\theta)\} = \alpha$ and*

$$(3)\qquad | c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2) N^{-(i-1)/2}, \qquad p = 0, 1, \cdots, s + 1 - i,\ \theta_1 \leq \theta \leq \theta_2.$$

Now let

$$(4)\qquad H_k(\theta) = N^{k/2} E(\hat{\theta} - \theta)^k = N^{k/2} \int_{-\infty}^{\infty} (\hat{\theta} - \theta)^k g_2(\hat{\theta} \mid \theta)\, d\hat{\theta} \quad \text{and}$$

$$(5)\qquad G_{pq}(\theta) = \frac{\partial^{p+q}}{\partial \hat{\theta}^p \partial t^q} g_1(t \mid \hat{\theta}, \theta) \mid_{t=c_1(\theta),\, \hat{\theta}=\theta}.$$

We have

THEOREM 2. *If*

(i) $\Pr\{t \leq c_1(\theta)\} = \alpha$, $0 < \alpha < 1$, *and* $| c_1^{(p)}(\theta) | \leq M(\theta_1, \theta_2)$, $\theta_1 \leq \theta \leq \theta_2$, $p = 0, 1, \cdots, s$;

(ii) $\delta = \delta(N) = O(1)$ *is a function of N such that*

$$\int_{|\hat{\theta}-\theta| \geq \delta} d\hat{\theta}\, | \hat{\theta} - \theta |^k g_2(\hat{\theta} \mid \theta) \leq M(\theta_1, \theta_2) N^{-s/2},\ \theta_1 \leq \theta \leq \theta_2,\ k = 0, 1, \cdots, s;$$

(iii) $\left| \frac{\partial^{p+q}}{\partial t^p \partial \hat{\theta}^q} g_1(t \mid \hat{\theta}, \theta) \right| \leq M(\theta_1, \theta_2)$, $p + q = s$, $| t - c_1(\theta) | < \rho, | \hat{\theta} - \theta | \leq \delta$,

where

$$\rho = \operatorname*{Max}_{|\hat{\theta}-\theta|\leq\delta} |\, c_1(\hat{\theta}) - c_1(\theta)\,| + N^{-(1/2)+\eta}, \quad \eta > 0, \quad \theta_1 \leq \theta \leq \theta_2\,;$$

(iv) $|\, H_k^{(p)}(\theta)\,| \leq M(\theta_1, \theta_2)$ *for* $p = 0, 1, \cdots, s-k, k = 1, \cdots, s,$ $\theta_1 \leq \theta \leq \theta_2$;

(v) $|\, G_{pq}^{(l)}(\theta)\,| \leq M(\theta_1, \theta_2)$ *for* $l = 0, 1, \cdots, s-p-q,$ $p+q \leq s-1$;

(vi) $\varphi_m(\hat{\theta})$ *is an admissible function of order* $m \leq s$,

then

$$\Pr\{t \leq \varphi_m(\hat{\theta})\} = \alpha + r_{m1}(\theta)N^{-1/2} + \cdots + r_{ms}(\theta)N^{-s/2} \tag{6}$$

where

$$|\, r_{mj}^{(p)}(\theta)\,| \leq M(\theta_1, \theta_2) \quad \text{for} \quad p = 0, 1, \cdots, s-j, \quad j \leq s, \quad \theta_1 \leq \theta \leq \theta_2\,.$$

PROOF. Expand $g_1(t \mid \hat{\theta}, \theta)$ in a Taylor Expansion about $t = c_1(\theta)$, $\hat{\theta} = \theta$, with remainder terms of order s in $t - c_1(\theta)$, $\hat{\theta} - \theta$, and expand $c_i(\theta)$ about $\hat{\theta} = \theta$ where the remainder term is of order $s + 1 - i$. Then for $|\, \hat{\theta} - \theta\,| \leq \delta$, we have

$$\int_{c_1(\theta)}^{\varphi_m(\hat{\theta})} g_1(t \mid \hat{\theta}, \theta)\, dt = P\{(\hat{\theta}-\theta)^j, c_i^{(p)}(\theta), G_{pq}\}_{-}^{+} + RN^{-s/2}, \tag{7}$$

where P is a polynomial and $|\, R\,| \leq M(\theta_1, \theta_2)\sum_{i=0}^{s} (\hat{\theta}-\theta)^{s-i}N^{-s/2}$ for $|\, \hat{\theta}-\theta\,| \leq \delta$. Integrating over $|\, \hat{\theta} - \theta\,| \leq \delta$, we use conditions (ii), (iv) and (v) and the theorem follows. By a similar argument we have

THEOREM 3. *If*

(i) *the conditions of Theorem 2 hold for each* (θ_1, θ_2) *so that*

$$-\infty \leq \beta_1 < \theta_1 < \theta_2 < \beta_2 \leq \infty$$

and

(ii) $g_1(c_1(\theta) \mid \theta, \theta) > (1/M(\theta_1, \theta_2)) > 0, \qquad \theta_1 \leq \theta \leq \theta_2\,,$

then the sequence

$$\begin{aligned} \varphi_1(\hat{\theta}) &= c_1(\hat{\theta}); \\ \varphi_2(\hat{\theta}) &= c_1(\hat{\theta}) - r_{1,1}(\hat{\theta})N^{-1/2}; \\ \varphi_m(\hat{\theta}) &= \varphi_{m-1}(\hat{\theta}) - \frac{r_{m-1,m-1}(\hat{\theta})N^{-(m-1)/2}}{g_1(c_1(\hat{\theta}) \mid \hat{\theta}, \hat{\theta})}, \qquad m \leq s, \end{aligned} \tag{8}$$

is a sequence of admissible functions such that

$$\Pr\{t \leq \varphi_m(\hat{\theta})\} = \alpha + R(\theta)N^{-m/2}, \qquad m \leq s, \tag{a}$$

where $|R(\theta)| \leq M(\theta_1, \theta_2)$ *for* $\beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2$.

These theorems permit us to obtain and to calculate critical regions whose size is asymptotically close to α.

In Theorem 2, condition (ii) was much stronger than necessary. It may be relaxed if we define

$$H_k(\theta) = \int_{|\hat{\theta}-\theta|\leq\delta} N^{k/2} g_2(\hat{\theta} \mid \theta)(\hat{\theta} - \theta)^k \, d\theta,$$

where

$$\Pr\{|\hat{\theta} - \theta| \geq \delta\} \leq M(\theta_1, \theta_2) N^{-s/2}, \qquad \delta = \delta(N) = O(1).$$

However this may complicate the calculations.

The symmetric case arises when the first moment almost vanishes, i.e.

$$|H_1^{(p)}(\theta)| \leq M(\theta_1, \theta_2) N^{-1/2}, \quad p = 0, 1, \cdots, s-1, \quad \theta_1 \leq \theta \leq \theta_2. \tag{10}$$

In this case we have instead of the sequence given in Theorem 3, the sequence

$$\begin{aligned}
\varphi_1(\hat{\theta}) &= c_1(\hat{\theta}); \\
\varphi_2(\hat{\theta}) &= c_1(\hat{\theta}) - \frac{r_{1,2}(\hat{\theta}) N^{-1} + r_{1,3}(\hat{\theta}) N^{-3/2}}{g_1(c_1(\hat{\theta}) \mid \hat{\theta}, \hat{\theta})}; \\
\varphi_m(\hat{\theta}) &= \varphi_{m-1}(\hat{\theta}) - \frac{r_{m-1,2m-2}(\hat{\theta}) N^{-(m-1)} + r_{m-1,2m-1}(\hat{\theta}) N^{-(2m-1/2)}}{g_1(c_1(\hat{\theta}) \mid \hat{\theta}, \hat{\theta})},
\end{aligned} \tag{11}$$

which is a sequence of admissible functions such that

$$\begin{aligned}
\Pr\{t \leq \varphi_m(\hat{\theta})\} &= \alpha + r_{m,2m}(\theta) N^{-m} + \cdots + r_{m,s}(\theta) N^{-s/2} \\
|r_{m,n}^{(p)}(\theta)| &\leq M(\theta_1, \theta_2) \qquad \theta_1 \leq \theta \leq \theta_2 \qquad p = 0, 1, \cdots, s - n.
\end{aligned}$$

4. An example. The following example previously treated by Welch from a different point of view will furnish an illustration of the applicability of the theorems to the case where θ is a k dimensional parameter. It will also serve as an example of an extended type of symmetry. That is, it has the property that $|H_{2k+1}^{(p)}(\theta)| \leq M(\theta_1, \theta_2) N^{-1/2}$, and hence, in the sequence (11), the $r_{m,2m+1}(\theta)$ terms effectively vanish thereby simplifying the calculations considerably

We suppose that t is a normally distributed variable with mean μ and variance $\sigma^2 = \lambda_1\sigma_1^2 + \cdots + \lambda_k\sigma_k^2$ where the λ_i are known positive constants, the σ_i^2 are unknown parameters each of which is independently estimated by s_i^2 where $N_i s_i^2/\sigma_i^2$ has the χ^2 distribution with N_i degrees of freedom. It is desired to test the hypothesis that $\mu = 0$ so that the probability of rejecting the hypothesis if it is true should equal α. Under the hypothesis the joint density distribution of $t, s_1^2, \cdots s_k^2$, is given by

$$g(t, s_1^2, \cdots, s_k^2/\sigma_1^2, \cdots, \sigma_k^2) = \frac{e^{-t^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^{k} g(s_i^2 \mid \sigma_i^2, N_i), \tag{12}$$

where the moments of $s_i^2 - \sigma_i^2 = \hat{\theta}_i - \theta_i$ are given by the coefficients of $u^k/k!$ in the expansion about $u = 0$ of $e^{-u\sigma_i^2}(1 - (2u\sigma_i^2/N_i))^{-N_i/2}$:

$$H_1(\sigma_i^2) = 0,$$
$$H_2(\sigma_i^2) = 2\sigma_i^4;$$
$$H_3(\sigma_i^2) = \tfrac{4}{3}\sigma_i^6 N_i^{-1/2};$$
$$H_4(\sigma_i^2) = (\tfrac{1}{6}\tfrac{1}{1} + 2N_i^{-1})\sigma_i^8.$$

We define $c_1(\theta)$ by $\Pr\{t \leq c_1(\theta)\} = \alpha$ where

$$\theta = (\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2) \quad \text{and} \quad \hat{\theta} = (s_1^2, s_2^2, \cdots, s_k^2),$$
$$c_1(\theta) = c_1\sigma.$$

Now $\alpha_2(\theta) - \alpha = \Pr\{c_1(\theta) \leq t \leq c_1(\hat{\theta})\}$ may be computed within terms of order N_i^{-2} by expanding

$$c_1(\hat{\theta}) \approx c_1\sigma + c_1 \sum \frac{1}{2\sigma}\lambda_i(s_i^2 - \sigma_i^2) - c_1 \sum \frac{1}{8\sigma^3}\lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2)$$
$$- \frac{1}{\sqrt{2\pi\sigma^2}}e^{-t^2/2\sigma^2} \approx \frac{1}{\sqrt{2\pi\sigma^2}}e^{-c_1^2/2}\{1 + (t - c_1\sigma)(-c_1/\sigma)\},$$

whence

$$\alpha_2(\theta) - \alpha \approx \int_0^\infty \cdots \int_0^\infty ds_1^2 \cdots ds_k^2 \prod_{i=1}^k g(s_i^2 \mid \sigma_i^2; N_i) \left\{\frac{e^{-c_1^2/2}}{\sqrt{2\pi\sigma^2}}\right\}\left\{\frac{c_1}{2\sigma}\sum \lambda_i(s_i^2 - \sigma_i^2)\right.$$
$$\left. - \frac{c_1}{8\sigma^3}\sum \lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2) - \frac{c_1}{2\sigma}\left(\frac{c_1^2}{4\sigma^2}\sum \lambda_i\lambda_j(s_i^2 - \sigma_i^2)(s_j^2 - \sigma_j^2)\right)\right\}$$
$$= -\frac{e^{-c_1^2/2}}{\sqrt{2\pi\sigma^2}}\left\{\frac{c_1 + c_1^3}{8\sigma^3}\right\}\{\sum 2\lambda_i^2\sigma_i^4 N_i^{-1}\} + O(\sum N_i^{-2}).$$

Thus

$$c_2(\theta) = \frac{c_1 + c_1^3}{4\sigma^3}\sum \lambda_i^2\sigma_i^4 N_i^{-1}$$

and

$$\alpha_3(\theta) = \Pr\{t \leq c_1 s + \frac{c_1 + c_1^3}{4s^3}\sum \lambda_i^2 s_i^4 N_i^{-1}\} = \alpha + O(\sum N_i^{-2}),$$

where

$$s^2 = \sum \lambda_i s_i^2.$$

Further approximations become somewhat complex and should be carried out in a systematic fashion.

5. Remarks. The range of application in practical statistical problems of the theorems of section 2 may be somewhat more limited than that of the original

method proposed by Wald. Concerning the original method, the following theorems have been established.

THEOREM 4. *If*

(i) $\Pr\{t \leq c_1(\theta)\} = \alpha$, $0 < \alpha < 1$, *where* $|\, c_1^{(p)}(\theta)\,| \leq M(\theta_1, \theta_2)$, $\theta_1 \leq \theta \leq \theta_2$, $p = 0, 1, \cdots, s$;

(ii) $\left|\, \frac{\partial^{i+j} g(t, \hat{\theta} + \Delta \mid \theta + \Delta)}{\partial t^i \partial \Delta^j} \,\right| \leq G(\hat{\theta}, \theta)$ *for* $i + j \leq s - 1$, $C_1 \leq t \leq C_2$; $\theta_1 \leq \theta$, $\theta + \Delta \leq \theta_2$, $|\, \Delta \,| \leq \Delta'$, *where* $G(\hat{\theta}, \theta)$ *depends on* $C_1, C_2, \theta_1, \theta_2, N$, *and is integrable in* $\hat{\theta}$ *over* $(-\infty, \infty)$;

(iii) $\left|\, \frac{\partial^{i+j} g(t, \hat{\theta} \mid \theta)}{\partial t^i \partial \Delta^j} \,\right| \leq L(\hat{\theta}, \theta)$, $i + j \leq s - 1$, $C_1 \leq t \leq C_2$, $\theta_1 \leq \theta \leq \theta_2$, *where* $\int_{-\infty}^{\infty} L(\hat{\theta}, \theta)\,|\, \hat{\theta} - \theta \,|^k\, d\hat{\theta} \leq M(\theta_1, \theta_2, C_1, C_2) N^{-k/2}$, $k = 0, 1$,

(iv) $0 < A(C_1, C_2, \theta_1, \theta_2) < A(t) \leq g_3(t \mid \theta) \leq B(t) < B(C_1, C_2, \theta_1, \theta_2) < \infty$; $\theta_1 \leq \theta \leq \theta_2$, $C_1 \leq t \leq C_2$, $\int_{-\infty}^{\infty} B(t)\, dt < M(\theta_1, \theta_2)$;

(v) $g(t, \hat{\theta} \mid \theta) > 0$,

then a sequence $c_1^*(\hat{\theta}), c_2(\theta), c_2^*(\hat{\theta}), \cdots, c_s^*(\hat{\theta})$, *exists where* $c_m(\theta)$ *is uniquely defined in* (θ_1, θ_2) *by* $\Pr\{t - c_1^*(\hat{\theta}) - \cdots - c_{m-1}^*(\hat{\theta}) \leq c_m(\theta)\} = \alpha$, *and*

$$|c_m^{(p)}(\theta)| \leq M(\theta_1, \theta_2) N^{-(m-1)/2}\ p = 0, 1, \cdots, s - m + 1,\ \theta_1 \leq \theta \leq \theta_2$$

and $c_m^*(\theta)$ *is any function so that*

$$|c_m^{*(p)}(\theta) - c_m^{(p)}(\theta)\,| \leq MN^{-m/2} \text{ for } \theta_1 \leq \theta \leq \theta_2,\ p = 0, 1, \cdots, s - m,$$

and

$$|\, c_m^{*(p)}(\hat{\theta})\,| \leq M(\theta_1, \theta_2) N^{-(m-1)/2} \quad -\infty < \hat{\theta} < \infty,\ p = 0, 1, \cdots, s - m + 1.$$

Finally for $c_m^*(\theta)$ arbitrary within the above conditions,

$$|\Pr\{t - c_1^*(\hat{\theta}) - \cdots - c_s^*(\hat{\theta}) \leq 0\} - \alpha\,| \leq M(\theta_1, \theta_2) N^{-s/2} \quad \text{for} \quad \theta_1 \leq \theta \leq \theta_2.$$

The conditions on the derivatives with respect to Δ are natural because the intuitive approach to the method seems to hinge on the assumption that $g(t, \hat{\theta} + \Delta \mid \theta + \Delta)$ changes gradually with respect to Δ "independent" of the value of N. This would not be true of $g(t, \hat{\theta} \mid \theta + \Delta)$ for large N.

The $c_i^*(\theta)$ where introduced in Theorem 4 because in practical examples it is usually found too difficult to compute $c_i(\theta)$ efficiently. On the other hand there are many alternative ways of obtaining functions with the properties

of the $c_i^*(\theta)$. The $c_2(\theta)$, $c_3(\theta)$ etc. mentioned in Theorems 1, 2, 3 play the role of the $c_i^*(\theta)$ in Theorem 4 with the exception of the condition on $c_i^*(\theta)$ for outside (θ_1, θ_2) The exception is due to the fact that the Theorems 1, 2, 3 correspond to the "infinite case." Theorem 4 is applicable to those cases where one is willing to assume that θ lies in (θ_1, θ_2). It often happens that there is no such reason or that the conditions of the theorem hold only for every closed proper subinterval of (β_1, β_2) but not for $\beta_1 \leq \theta \leq \beta_2$ itself. In these cases we may apply

THEOREM 5. *If*

(i) *all of the conditions of Theorem 4 apply to every finite proper closed subinterval* (θ_1, θ_2) *of* (β_1, β_2) *where* (β_1, β_2) *may be an infinite interval*;

(ii) $\Pr\{|\hat{\theta} - \theta| \geq \delta(N)\} \leq M(\theta_1, \theta_2)N^{-s/2}$ *for* $\beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2$, *where* $\delta(N) = 0(1)$ *unless* β_1 *or* β_2 *is finite, in which case* $\delta(N) = o(1)$, *then a sequence* $c_1^*(\hat{\theta})$, $c_2(\theta)$, $c_2^*(\hat{\theta})$, $\cdots$, $c_s^*(\hat{\theta})$, *exists, where* $c_m(\theta)$ *is uniquely defined in* (β_1, β_2) *by* $Pr\{t - c_1^*(\hat{\theta}) - c_2^*(\hat{\theta}) - \cdots - c_{m-1}^*(\hat{\theta}) \leq c_m(\theta)\} = \alpha$, *so that for every* (θ_1, θ_2),

$$|c_m^{*(p)}(\theta)| \leq M(\theta_1, \theta_2)N^{-(m-1)/2} \text{ if } \beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2,$$
$$p = 0, 1, \cdots, s - m + 1'$$

and for $c_m^*(\theta)$ *arbitrary within the above conditions*

$$|\Pr\{t \leq c_1^*(\hat{\theta}) + \cdots + c_m^*(\hat{\theta})\} - \alpha| \leq M(\theta_1, \theta_2)N^{-m/2}$$
$$\text{if} \quad \beta_1 < \theta_1 \leq \theta \leq \theta_2 < \beta_2, m \leq s.$$

Essentially this theorem can be proved by reference to the proof of Theorem 4 applied to the function

$$g^*(t, \hat{\theta} \mid \theta) = g(t, \hat{\theta} \mid \theta) \quad \text{for} \quad |\hat{\theta} - \theta| \leq \delta;$$
$$= 0 \qquad\qquad |\hat{\theta} - \theta| > \delta.$$

Some of the conditions in Theorems 4 and 5 are stronger than necessary. For example $g > 0$ may be replaced by a weaker condition where g is positive in a region about $t = c_1(\theta)$. On the other hand the condition $\Pr\{|\hat{\theta} - \theta| \geq \delta\} \leq MN^{-s/2}$ in Theorem 5 is necessary to the argument used in the proof. It is easy to construct trivial examples where the results of this theorem apply although this condition is not satisfied. However an example has also been constructed where all the conditions of Theorem 5 hold except for this condition and the method of Wald fails to give the results.

These theorems are very easily extended to the k-dimensional parameter case by replacing the conditions on the derivatives with respect to Δ by the same order mixed derivatives with respect to $\Delta_1, \Delta_2, \cdots, \Delta_k$ of

$$g(t, \hat{\theta}_1 + \Delta_1, \hat{\theta}_2 + \Delta_2, \cdots, \hat{\theta}_k + \Delta_k \mid \theta_1 + \Delta_1, \cdots, \theta_k + \Delta_k).$$

The symmetric case arises when the distribution of $\hat{\theta}$ is almost symmetric about θ. More exactly we have

THEOREM 6. *If*

(i) *All the conditions of Theorem* 4 *hold and* $L(\hat{\theta}, \theta)$ *has the additional property that*

$$\int_{-\infty}^{\infty} (\hat{\theta} - \theta)^2 L(\hat{\theta}, \theta)\, d\hat{\theta} < M(\theta_1, \theta_2)N^{-1}, \qquad \theta_1 \leq \theta \leq \theta_2,$$

and

(ii)
$$\left| \frac{\partial g^{i+j}}{\partial \Delta^i \partial t^j}(t, \hat{\theta} \mid \theta) - \frac{\partial g^{i+j}}{\partial \Delta^i \partial t^j}(t, 2\theta - \hat{\theta} \mid \theta) \right| < L(\hat{\theta}, \theta) \mid \hat{\theta} - \theta \mid,$$

$$C_1 \leq t \leq C_2, \qquad \theta_1 \leq \theta \leq \theta_2, \qquad i + j \leq s - 1,$$

then it is possible to construct a sequence $c_1^*(\hat{\theta}), c_2(\theta), \cdots, c_r^*(\hat{\theta})$, *as in Theorem* 4 *so that*

$$\mid c_m^{(p)}(\theta) \mid \leq M(\theta_1, \theta_2)N^{-(m-1)},$$

$$p = 0, 1, \cdots, s - 2m + 2, \theta_1 \leq \theta \leq \theta_2;$$

$$\mid c_m^{*(p)}(\theta) - c_m^{(p)}(\theta) \mid \leq M(\theta_1, \theta_2)N^{-m},$$

$$p = 0, 1, \cdots, s - 2m + 1, \theta_1 \leq \theta \leq \theta_2;$$

$$\mid c_m^{*(p)}(\hat{\theta}) \mid \leq M(\theta_1, \theta_2)N^{-(m-1)},$$

$$p = 0, 1, \cdots, s - 2m + 2, -\infty < \hat{\theta} < \infty;$$

and

$$\mid \Pr\{t \leq c_1^*(\hat{\theta}) + \cdots + c_r^*(\hat{\theta})\} - \alpha \mid \leq M(\theta_1, \theta_2)N^{-s/2},$$

$$\theta_1 \leq \theta \leq \theta_2, r = \left[\frac{s+1}{2}\right].$$

Theorem 5 can also be extended to the symmetric case

It is often possible in the theory of statistics to obtain an asymptotic expansion of the distribution of t, $\hat{\theta}$. The treatment of such cases is often very simple because of the prominent role played by the normal distribution in such asymptotic expansions. Suppose that

$$g(t, \hat{\theta} \mid \theta) = \sqrt{N}\gamma(t, \psi \mid \theta),$$

where $\psi = \sqrt{N}(\hat{\theta} - \theta)$; γ = density distribution of (t, ψ);

$$\gamma(t, \psi \mid \theta) = \gamma_0(t, \psi \mid \theta) + N^{-1/2}\gamma_1(t, \psi \mid \theta) + \cdots + N^{-(s-1)/2}\gamma_{s-1}(t, \psi \mid \theta) + \rho(t, \psi \mid \theta)N^{-s/2},$$

$\gamma_0, \gamma_1, \cdots, \gamma_{s-1}$ are independent of N;

$$\int\int \mid \rho \mid d\psi\, dt \leq M(\theta_1, \theta_2), \qquad \theta_1 \leq \theta \leq \theta_2;$$

$$\int\int \mid \gamma_i \mid d\psi\, dt \leq M(\theta_1, \theta_2), \qquad \theta_1 \leq \theta \leq \theta_2.$$

Correspondingly we have

$$g(t, \hat{\theta} \mid \theta) = g_0(t, \hat{\theta} \mid \theta) + N^{-1/2} g_1(t, \hat{\theta} \mid \theta) + \cdots + N^{-(s-1)/2} g_{s-1}(t, \hat{\theta} \mid \theta) + r(t, \hat{\theta} \mid \theta) N^{-s/2},$$

where

$$g_i(t, \hat{\theta} \mid \theta) = \sqrt{N}\, \gamma_i(t, \psi \mid \theta), \qquad r(t, \hat{\theta} \mid \theta) = \sqrt{N} \rho(t, \psi \mid \theta).$$

Then if we define $c_1(\theta)$ by $\int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\theta)} dt g_0 = \alpha$,

$c_m(\theta)$ by $\int_{-\infty}^{\infty} d\hat{\theta} \int_{c_1(\hat{\theta})+\cdots+c_{m-1}(\hat{\theta})}^{c_1(\theta)+\cdots+c_{m-1}(\hat{\theta})+c_m(\theta)} dt g_0$

$$= \alpha - \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\hat{\theta})+\cdots+c_{m-1}(\theta)} dt [g_0 + g_1 N^{-1/2} + \cdots + g_{m-1} N^{-(m-1)/2}],$$

or by

$$c_m(\theta) \int_{-\infty}^{\infty} d\hat{\theta} g_0(c_1(\theta), \hat{\theta} \mid \theta)$$

$$= \alpha - \int_{-\infty}^{\infty} d\hat{\theta} \int_{-\infty}^{c_1(\hat{\theta})+\cdots+c_{m-1}(\hat{\theta})} dt [g_0 + g_1 N^{-1/2} + \cdots + g_{m-1} N^{-((m-1)/2)}],$$

we obtain

$$| \Pr\{t \leq c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})\} - \alpha | \leq M(\theta_1, \theta_2) N^{-s/2},$$

if g obeys the conditions of Theorem 4 except that we need only $s - i + 1$ derivatives for $g_i(t, \hat{\theta} \mid \theta)$. The above definitions of $c_m(\theta)$ correspond to the $c_m^*(\theta)$ in Theorem 4. Analogues of Theorems 5 and 6 also apply to the asymptotic case.

REFERENCE

[1] B L. Welch, "On the studentization of several variances," *Annals of Math. Stat.*, Vol. 18 (1947), p. 118.

SOME LOW MOMENTS OF ORDER STATISTICS

By H. J. Godwin

University College of Swansea, Wales

1. Introduction. In a paper on order statistics from several populations [1], there were given, among other results, the means, variances, covariances, and correlations of order statistics in samples of ten or less from a normal population. These were obtained by numerical integration, and on account of the difficulties arising therefrom, some results were given to only two decimal places. More recently, Jones [3] has shown that some of the integrals, for sample sizes not greater than four, can be evaluated explicitly.

In this note these results are supplemented in two ways. For a paper which the author has recently submitted to *Biometrika* integrals were evaluated which can be used to give some of the results in [1] to more places of decimals. It is also shown that the table of explicit values can be extended.

2. Approximate values. Let the population studied be normal with mean zero and variance unity, and let the members of a sample of n be $x(1 \mid n) \geq x(2 \mid n) \geq \cdots \geq x(n \mid n)$. The integrals available are

$$\psi(i) = \int_{-\infty}^{\infty} F^i(x)(1 - F(x))^i \, dx (1 \leq i \leq 5), \quad \text{and}$$

$$\psi(i, j) = \int_{-\infty}^{\infty} F^i(x) \int_x^{\infty} (1 - F(y)) \, dx \, dy (1 \leq i, j; i + j \leq 10),$$

where

$$F(x) = \int_{-\infty}^{x} f(t) \, dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dt.$$

These were evaluated to ten places of decimals, the last place possibly being in error by one or two units.

For the purpose in hand we define also

$$\alpha(i, j) = \int_{-\infty}^{\infty} x F^i(x)(1 - F(x))^j \, dx = -\alpha(j, i),$$

and

$$\beta(i, j) = \int_{-\infty}^{\infty} x^2 f(x) F^i(x)(1 - F(x))^j \, dx = \beta(j, i).$$

Now, on integrating by parts, we have

$$\int_x^{\infty} (1 - F(y)) \, dy = -x(1 - F(x)) + \int_x^{\infty} y f(y) \, dy,$$

and for $f(x)$ as defined above (so that in what follows we restrict ourselves to the normal distribution only), the second integral is $f(x)$. Hence $\psi(i, 1) + \alpha(i, 1) = 1/(i + 1)$ and we can construct a table of α's by using also the relation

$$\alpha(i, j) - \alpha(i + 1, j) = \alpha(i, j + 1).$$

Again, on integrating by parts, we have

$$\beta(i, i) = \int_{-\infty}^{\infty} \frac{F^{i+1}(x)}{i+1} \{ix^2 f(x)(1 - F(x))^{i-1} - 2x(1 - F(x))^i\}\, dx (i > 0)$$

$$= \frac{i}{i+1} \{\tfrac{1}{2}\beta(i - 1, i - 1) - \beta(i, i)\} - \frac{2}{i+1} \alpha(i + 1, i),$$

using the fact that, in this particular case, $2F - 1$ is an odd function and $F(1 - F)$ an even function of x.

Hence $\beta(i, i) = \dfrac{i}{2(2i+1)} \beta(i - 1, i - 1) - \dfrac{2}{2i+1} \alpha(i + 1, i)$, and using $\beta(i, j) - \beta(i + 1, j) = \beta(i, j + 1)$ we can find the β's.

Finally we put $\gamma(i, j) = \dfrac{i\beta(i - 1, j) + \alpha(i, j) - \psi(i, j)}{ij}$ which can be shown by an integration to be equal in this case to $\gamma(j, i)$.

Now

$$(1) \qquad E(x(i \mid n) - x(i + 1 \mid n)) = {}^nC_i \int_{-\infty}^{\infty} F^{n-i}(x)(1 - F(x))^i\, dx,$$

as was proved by Irwin [2]. By the symmetry here this integral is the same if $i, n - i$ are interchanged, and since $F^a(1 - F)^b + F^b(1 - F)^a$ is a polynomial in $F(1 - F)$ (as may be seen by putting $F = \frac{1}{2} + G$) the integrals (1) can be expressed in terms of the $\psi(i)$. Using the fact that the expected value of the median is zero the $E(x(i \mid n))$ follow.

The frequency function of $x = x(i \mid n)$ is

$$\frac{n!}{(i - 1)!(n - i)!} f(x)(1 - F(x))^{i-1} F^{n-i}(x),$$

and so

$$(2) \qquad E(x(i \mid n))^2 = i\, {}^nC_i\, \beta(i - 1, n - i).$$

The joint frequency function of $x_i = x(i \mid n)$ and $x_j = x(j \mid n)$ is

$$\frac{n!}{(i - 1)!(j - i - 1)!(n - j)!} f(x_i) f(x_j)(1 - F(x_i))^{i-1} (F(x_i) - F(x_j))^{j-i-1} F^{n-j}(x_j)$$

(taking $j > i$), and to find $E(x_i x_j)$ we multiply by $x_i x_j$ and integrate, x_j going

from $-\infty$ to ∞, and x_i from x_j to ∞ On expanding $(1-F(x_j)-(1-F(x_i)))^{j-i-1}$ by the multinomial theorem a typical term is

$$\int_{-\infty}^{\infty}\int_{x_j}^{\infty} x_i x_j f(x_i)f(x_j)(1 - F(x_i))^{i-1+r} F^{n-j+s}(x_j)\, dx_j\, dx_i . \tag{3}$$

TABLE 1

Means and standard deviations

Statistic	Mean	Standard Deviation	Statistic	Mean	Standard Deviation
$x(1\|2)$	.5641896	.8256453	$x(1\|8)$	1.4236003	.6106530
$x(1\|3)$	.8462844	.7479754	$x(2\|8)$	8522249	4892862
$x(2\|3)$	0	.6698292	$x(3\|8)$	.4728225	.4480723
$x(1\|4)$	1.0293754	7012241	$x(4\|8)$	.1525144	.4326503
$x(2\|4)$	.2970114	.6003793	$x(1\|9)$	1.4850132	.5977903
$x(1\|5)$	1.1629645	.6689799	$x(2\|9)$	.9322975	.4750755
$x(2\|5)$	.4950190	.5581388	$x(3\|9)$	.5719708	.4317205
$x(3\|5)$	0	.5355685	$x(4\|9)$	.2745259	.4129877
$x(1\|6)$	1.2672064	.6449241	$x(5\|9)$	0	.4075553
$x(2\|6)$	.6417550	.5287511	$x(1\|10)$	1.5387527	.5868083
$x(3\|6)$	.2015468	.4961981	$x(2\|10)$	1.0013571	.4631674
$x(1\|7)$	1.3521784	.6260334	$x(3\|10)$	.6560591	.4183339
$x(2\|7)$	.7573743	.5066882	$x(4\|10)$	.3757647	.3974153
$x(3\|7)$	.3527070	.4687447	$x(5\|10)$	.1226678	.3886565
$x(4\|7)$	0	.4587449			

We integrate by parts with respect to x_i and then with respect to x_j : the integral (3) is then seen to be $\gamma(i + r, n - j + s + 1)$, and

$$E(x_i x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \sum_{r=0}^{j-i-1} \sum_{s=0}^{j-i-1-r} \cdot \frac{(-1)^{r+s}(j-i-1)!}{r!s!(j-i-1-r-s)!} \gamma(i + r, n - j + s + 1). \tag{4}$$

Using (1), (2) and (4), the values in Tables 1, 2, and 3 are obtained. The values are estimated to be correct, except for sample sizes 9 and 10, for which there may be errors of one or two units in the last place given. Missing values are filled in by considerations of symmetry.

3. Exact values. All the integrals occurring for $\psi(i)$ or $\psi(i, j)$ can, by suitable transformations, and the integration of one variable over the range $-\infty$ to ∞,

TABLE 2

Variances and covariances

n	i	j									
		1	2	3	4	5	6	7	8	9	10
2	1	.68169	.31831								
3	1	.55947	.27566	.16487							
	2		.44867								
4	1	.49172	.24559	.15801	.10468						
	2		.36046	.23594							
5	1	.44753	.22433	.14815	.10577	.07422					
	2		.31152	.20844	.14994						
	3			.28683							
6	1	.41593	.20850	.13944	.10243	.07736	.05634				
	2		.27958	.18899	.13966	.10591					
	3			.24621	18327						
7	1	.39192	.19620	.13212	.09849	.07656	.05992	.04480			
	2		.25673	17448	.13073	.10196	.07998				
	3			.21972	.16556	.12960					
	4				.21045						
8	1	.37290	.18631	.12597	.09472	.07477	.06021	.04830	.03684		
	2		.23940	16320	.12326	.09757	.07872	.06325			
	3			20077	.15236	.12096	.09782				
	4				.18719	14918					
9	1	.35735	.17814	.12075	.09131	.07274	.05948	.04908	.04009	.03106	
	2		.22570	.15412	.11701	.09345	.07655	.06324	.05171		
	3			.18638	.14208	.11377	.09336	.07723			
	4				.17056	.13699	.11267				
	5					.16610					
10	1	.34434	.17126	.11626	08825	.07074	.05840	.04892	.04108	.03404	.02675
	2		.21452	.14662	.11170	.08974	.07420	.06222	.05232	.04336	
	3			.17500	.13380	.10774	.08923	.07492	.06302		
	4				.15794	.12751	.10579	.08895			
	5					.15105	.12560				

TABLE 3

Correlations between order statistics

n	i	j								
		2	3	4	5	6	7	8	9	10
2	1	.4669								
3	1	.5502	.2947							
4	1	.5834	.3753	.2129						
	2		.6546							
5	1	.6008	.4135	.2833	.1658					
	2		.6973	.4813						
6	1	.6114	.4357	.3201	.2269	.1355				
	2		.7203	.5323	.3788					
	3			.7444						
7	1	.6185	.4502	.3429	.2609	.1889	.1143			
	2		.7346	.5624	.4293	.3115				
	3			.7699	.5899					
8	1	.6236	.4604	.3585	.2830	.2200	.1617	0988		
	2		.7444	.5823	.4609	.3591	.2642			
	3			.7859	.6240	.4872				
	4				.7969					
9	1	.6273	.4679	.3699	.2986	.2409	.1902	.1412	.0869	
	2		.7514	.5964	.4827	.3902	.3083	:2291		
	3			.7969	.6466	.5236	4144			
	4				.8139	.6606				
10	1	.6301	.4736	.3784	.3102	.2561	.2098	.1674	.1252	.0777
	2		.7567	.6068	.4985	.4122	.3380	.2700	.2021	
	3			.8048	.6627	.5488	.4507	.3601		
	4				.8255	.6849	5632			
	5					.8315				

be represented as multiples of $\int_0^\infty \cdots \int e^{-Q}\, dx\, dy \cdots$, where Q is a positive-definite quadratic form in the variables of integration.

Now if Q is ax^2, the integral is $\frac{1}{2}\sqrt{\pi/a}$ (this is, in effect, stated by Jones). By elementary integration we have also that if $Q = ax^2 + 2hxy + by^2$, the integral is

$$\frac{1}{\sqrt{ab - h^2}}\left\{\frac{\pi}{2} - \arctan\frac{h}{\sqrt{ab - h^2}}\right\}$$

TABLE 4

Exact expected values

$x(1\|4)$:	$\sqrt{\pi}\,[(2/5)a$		$+ (2/5)c]$			
$x(2\|4)$:	$\sqrt{\pi}\,[(2/5)a$		$- (6/5)c]$			
$x(1\|5)$:	$\sqrt{\pi}\,[(1/3)a$		$+c]$			
$x(2\|5)$:	$\sqrt{\pi}\,[(2/3)a$		$-2c]$			
$x(3\|5)$:		0				
$x(1\|5)^2$:	1	$+b$		$+d$		
$x(2\|5)^2$:	1			$-4d$		
$x(3\|5)^2$:	1	$-2b$		$+6d$		
$x(1\|5)x(2\|5)$:		b		$+d$		
$x(1\|5)x(3\|5)$:	$2a$	$-2b$		$-2d$	$-f$	
$x(1\|5)x(4\|5)$:	$-2a$				$+3f$	
$x(1\|5)x(5\|5)$:					$-2f$	
$x(2\|5)x(3\|5)$:	$-2a$	$+3b$		$-d$	$+f$	
$x(2\|5)x(4\|5)$:	$4a$	$-4b$		$+4d$	$-4f$	
$x(1\|6)^2$:	1	$+b$		$+3d$		
$x(2\|6)^2$:	1	$+b$		$-9d$		
$x(3\|6)^2$:	1	$-2b$		$+6d$		
$x(1\|6)x(2\|6)$:		b		$+3d$		
$x(1\|6)x(3\|6)$:	$3a$	$-2b$	$+3c$	$-6d$	$-3f$	
$x(1\|6)x(4\|6)$:	$-3a$		$-9c$		$+9f$	
$x(1\|6)x(5\|6)$:			$12c$		$-6f$	
$x(1\|6)x(6\|6)$:			$-6c$			
$x(2\|6)x(3\|6)$:	$-3a$	$+4b$	$-3c$		$+3f$	
$x(2\|6)x(4\|6)$:	$9a$	$-6b$	$+9c$	$+6d$	$-15f$	
$x(2\|6)x(5\|6)$.	$-6a$		$-18c$		$+18f$	
$x(3\|6)x(4\|6)$:	$-6a$	$+6b$		$-6d$	$+6f$	

and if Q is $ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy$, the integral is

$$\frac{1}{4}\sqrt{\frac{\pi}{\Delta}}\left\{\frac{\pi}{2} + \arctan\frac{gh - af}{\sqrt{a\Delta}} + \arctan\frac{hf - bg}{\sqrt{b\Delta}} + \arctan\frac{fg - ch}{\sqrt{c\Delta}}\right\},$$

Where $\Delta = abc + 2fgh - af^2 - bg^2 - ch^2$.

The author has not succeeded in obtaining similar results with a higher number of variables—it is possible that elementary functions no longer suffice then

Using these results we can obtain exact expressions for $\psi(1)$, $\psi(2)$ and $\psi(i, j)$ for $1 \leq i, j; i + j \leq 6$, which give, in addition to Jones' results, the exact expected values in Table 4, wherein

$$a = 15/4\pi = 1.19366\ 20732,$$

$$b = 5\sqrt{3}/4\pi = .68916\ 11193,$$

$$c = (15/2\pi^2) \text{ arc sin } (1/3) = .25824\ 50843,$$

$$d = (5\sqrt{3}/2\pi^2) \text{ arc sin } \tfrac{1}{4} = .11085\ 93167,$$

$$f = (15/\pi^2) \text{ arc sin } (1/\sqrt{6}) = .63913\ 55493.$$

REFERENCES

[1] C. Hastings, Jr., F. Mosteller, J. W. Tukey and C. P. Winsor, "Low moments for small samples: a comparative study of order statistics," *Annals of Math. Stat.*, Vol. 18 (1947).

[2] J. O. Irwin, "The further theory of Francis Galton's individual-difference problem," *Biometrika*, Vol. 17 (1925).

[3] H. L. Jones, "Exact lower moments of order statistics in small samples from a normal distribution," *Annals of Math Stat.*, Vol. 19 (1948).

ON A THEOREM OF HSU AND ROBBINS

By P. Erdös

Syracuse University

Let $f_1(x), f_2(x), \cdots$ be an infinite sequence of measurable functions defined on a measure space X with measure m, $m(X) = 1$, all having the same distribution function $G(t) = m(x; f_k(x) \leq t)$. In a recent paper Hsu and Robbins[1] prove the following theorem: *Assume that*

$$\int_{-\infty}^{\infty} t\, dG(t) = 0, \tag{1}$$

$$\int_{-\infty}^{\infty} t^2\, dG(t) < \infty. \tag{2}$$

Denote by S_n *the set* $\left(x, \left|\sum_{k=1}^{n} f_k(x)\right| > n\right)$, *and put* $M_n = m(S_n)$. *Then* $\sum_{n=1}^{\infty} M_n$ *converges.*

It is clear that the same holds if $\left|\sum_{k=1}^{n} f_k(x)\right| > n$ is replaced by $\left|\sum_{k=1}^{n} f_k(x)\right| > c \cdot n$ (replace $f_k(x)$ by $c \cdot f_k(x)$)

It was conjectured that the conditions (1) and (2) are necessary for the convergence of $\sum_{n=1}^{\infty} M_n$. Dr. Chung pointed it out to me that in this form the conjecture is inaccurate; to see this it suffices to put $f_k(x) = \frac{1}{2}(1 + r_k(x))$ where $r_k(x)$ is the kth Rademacher function. Clearly $|f_k(x)| < 1$; thus $M_n = 0$, thus $\sum_{n=1}^{\infty} M_n$ converges, but $\int_{-\infty}^{\infty} t\, dG(t) \neq 0$. On the other hand we shall show in the present note that the conjecture of Hsu and Robbins is essentially correct. In fact we prove

Theorem I. *The necessary and sufficient condition for the convergence of* $\sum_{n=1}^{\infty} M_n$ *is that*

$$\left|\int_{-\infty}^{\infty} t\, dG(t)\right| < 1, \tag{1'}$$

and (2) *should hold.*

In proving the sufficiency of Theorem I, we can assume without loss of generality that (1) holds. It suffices to replace $f_k(x)$ by $(f_k(x) - C)$ where $C = \int_{-\infty}^{\infty} t dG(t)$. The following proof of the sufficiency of Theorem I (in other words essentially for the theorem of Hsu and Robbins) is simpler and quite different from theirs. Put

$$a_i = m(x; |f_k(x)| > 2^i), \tag{3}$$

[1] *Proc. Nat Acad. Sciences*, 1947, pp. 25–31.

since the f_k's all have the same distribution, a_i clearly does not depend on k. We evidently have

$$\sum_{i=0}^{\infty} 2^{2i-1} a_i \leq \sum_{i=0}^{\infty} 2^{2i}(a_i - a_{i+1}) \leq \int_{-\infty}^{\infty} t^2\, dG(t) \leq \sum_{i=0}^{\infty} 2^{2i+2}(a_i - a_{i+1}) \leq \sum_{i=0}^{\infty} 2^{2i+2} a_i.$$

Thus (2) is equivalent to

$$\sum_{i=0}^{\infty} 2^{2i} a_i < \infty. \tag{4}$$

Let $2^i \leq n < 2^{i+1}$. Put

$S_n^{(1)} = (x; |f_k(x)| > 2^{i-2}$, for at least one $k \leq n)$,

$S_n^{(2)} = (x; |f_{k_1}(x)| > n^{4/5}, |f_{k_2}(x)| > n^{4/5}$, for at least two $k_1 \leq n, k_2 \leq n)$,

$S_n^{(3)} = (x; \left|\sum_{k=1}^{n}{}' f_k'(x)\right| > 2^{i-2})$,

where the dash indicates that the k with $|f_k(x)| > n^{4/5}$ are omitted. We evidently have

$$S_n \subset S_n^{(1)} \cup S_n^{(2)} \cup S_n^{(3)}.$$

For if x is not in $S_n^{(1)} \cup S_n^{(2)} \cup S_n^{(3)}$, then clearly

$$\left|\sum_{k=1}^{n} f_k(x)\right| \leq 2^{i-2} + 2^{i-2} < n.$$

Thus to prove the convergence of $\sum_{n=1}^{\infty} M_n$ it will suffice to show that

$$\sum_{n=1}^{\infty} (m(S_n^{(1)}) + m(S_n^{(2)}) + m(S_n^{(3)})) < \infty. \tag{5}$$

From (3) we obtain that $m(S_n^{(1)}) \leq n \cdot a_{i-2} < 2^{i+1} \cdot a_{i-2}$. Thus from (4)

$$\sum_{n=1}^{\infty} m(S_n^{(1)}) = \sum_{i=0}^{\infty} \sum_{2^i \leq n < 2^{i+1}} m(S_n^{(1)}) < \sum_{i=0}^{\infty} 2^{2i+3} a_i < \infty. \tag{6}$$

From (4) we evidently have that for large u

$$m(x; |f_k(x)| > u) < 1/u^2.$$

Thus since the f's are independent and have the same distribution function it follows that for sufficiently large n,

$$m(S_n^{(2)}) \leq \sum_{1 \leq k_1 < k_2 \leq n} m(x; |f_{k_1}(x)| > n^{4/5}, |f_{k_2}(x)| > n^{4/5})$$

$$\leq \binom{n}{2} m(x; |f_1(x)| > n^{4/5}), m(x; |f_2(x)| > n^{4/5}) < n^2 \cdot n^{-16/5} = n^{-6/5}.$$

Hence

$$\sum_{n=1}^{\infty} m(S_n^{(2)}) < \infty. \tag{7}$$

Put

$$f_k^+(x) = \begin{cases} f_k(x) \text{ for } |f_k(x)| < n^{4/5}; \\ 0 \text{ otherwise.} \end{cases}$$

Clearly the $f_k^+(x)$ are independent and have the same distribution function $G^+(t)$. Put

$$\int_{-\infty}^{\infty} t\, dG^+(t) = \epsilon, \qquad g_k(x) = f_k^+(x) - \epsilon. \tag{8}$$

We have from (8) that $\int_X g_k(x)\, dm = 0$, and by (1) that $\epsilon \to 0$ as $n \to \infty$. We evidently have

$$\int_X \left(\sum_{k=1}^n g_k(x)\right)^4 dm = \int_X \sum_{k=1}^n g_k^4(x)\, dm + 6 \int_X \sum_{1 \le k < l \le n} g_k^2(x) \cdot g_l^2(x)\, dm.$$

Now since $\max |g_k(x)| < n^{4/5} + \epsilon$,

$$\int_X g_k^4(x)\, dm < (n^{4/5} + \epsilon)^2 \cdot \int_X g_k^2(x)\, dm < c_1 \cdot n^{8/5},$$

and

$$\int_X g_k^2(x) \cdot g_l^2(x)\, dm = \int_X g_k^2(x)\, dm \int_X g_l^2(x)\, dm < c_2.$$

Thus

$$\int_X \left(\sum_{k=1}^n g_k(x)\right)^4 dm < c_3\, n^{13/5}.$$

Hence

$$m\left(x; \left|\sum_{k=1}^n g_k(x)\right| > n/16\right) < c_4\, n^{-(7/5)}. \tag{9}$$

Thus from (8), (9), $|f_k^+(x)| < |g_k(x)| + 1/16$ (for $\epsilon < 1/16$) and $n/8 < 2^{i-2}$ we have

$$m\left(x; \left|\sum_{k=1}^n f'(x)\right| > 2^{i-2}\right) = m\left(x; \left|\sum_{k=1}^n f_k^+(x)\right| > 2^{i-2}\right)$$
$$< m\left(x; \left|\sum_{k=1}^n g_k(x)\right| > n/16\right) < c_4 n^{-(7/5)},$$

or

$$m(S_n^{(3)}) < c_4\, n^{-(7/5)}. \tag{10}$$

Thus finally from (6), (7) and (10) we obtain (5) and this completes the proof of the sufficiency of Theorem I.

Next we prove the necessity of Theorem I, in other words we shall show that if $\sum_{n=1}^{\infty} M_n$ converges then (1′) and (2) hold.

First we prove (2). The following proof was suggested by Dr. Chung, who simplified my original proof. By a simple rearrangement we see that (2) is equivalent to

$$\sum_{n=1}^{\infty} n \int_{|t|>cn} dG(t) < \infty \tag{11}$$

for any $c > 0$; while

$$\int_{-\infty}^{\infty} |t| \, dG(t) < \infty \tag{12}$$

is equivalent to

$$\sum_{n=1}^{\infty} \int_{|t|>cn} dG(t) < \infty \tag{13}$$

for any $c > 0$. Now we have clearly,

$$(x; |f_n(x)| > 2n) \subset S_{n-1} \cup S_n .$$

Hence

$$\sum_n \int_{|t|>2n} dG(t) \leqq \sum_n (m(S_{n-1}) + m(S_n)) < \infty.$$

Thus we obtain (12). Since the terms of this series is non-increasing it follows that

$$n \int_{|t|>2n} dG(t) \to 0. \tag{14}$$

Our assumption being that $\Sigma M_n < \infty$ we have $M_n \to 0$ as $n \to \infty$. It follows that there is a constant $\rho > 0$ independent of k and n such that

$$m\left(x; \left|\sum_{\substack{l=1 \\ l\neq k}}^{n} f_l(x)\right| < n\right) \geqq \rho.$$

Now, writing set intersections as products, we have

$$\bigcup_{k=1}^{n} (x; |f_k(x)| > 2n) \cdot \left(x; \left|\sum_{\substack{l=1 \\ l\neq k}}^{n} f_l(x)\right| < n\right) \subset S_n .$$

Writing this for a moment as

$$\bigcup_{k=1}^{n} (R_k T_k) \subset S_n ,$$

where $R_k = (x; |f_k(x)| > 2n)$ etc. and denoting by R' the complement of R, we have

$$M_n = m(S_n) \geqq m\left(\bigcup_{k=1}^{n} (R_k \cdot T_k)\right)$$

$$= m\left(\bigcup_{k=1}^{n} (R_1 T_1)' \cdots (R_{k-1} T_{k-1})' R_k T_k\right)$$

$$= \sum_{k=1}^{n} m((R_1 T_1)' \cdots (R_{k-1} T_{k-1})' R_k T_k)$$

$$\geqq \sum_{k=1}^{n} m(R_1' \cdots R_{k-1}' R_k T_k)$$

$$\geqq \sum_{k=1}^{n} \{m(R_k \cdot T_k) - m((R_1 \cup \cdots \cup R_{k-1}) R_k)\}$$

$$\geqq \sum_{k=1}^{n} \{m(T_k) - (k-1)m(R_1))m(R_k)\}$$

$$\geqq \sum_{k=1}^{n} \{\rho - nm(R_1)\}m(R_k) \geqq \sum_{k=1}^{m} (\rho - \sigma(1))m(R_k).$$

$$\geqq \rho' \sum_{k=1}^{n} m(R_k) = n\rho' \int_{|t|>2n} dG(t)$$

by (14) since $m(R_1) = \int_{|t|>2n} dG(t)$, $nm(R_1) \to 0$ as $n \to \infty$.

Thus

$$\sum_n n \int_{|t|>2n} dG(t) \leqq \frac{1}{\rho'} \sum_n M_n < \infty.$$

Hence we have (11), which is equivalent to (2). The proof of (1′) is quite easy. By virtue of (2) we can put

$$\int_{-\infty}^{\infty} tG(t) = C.$$

If $C > 1$, then it follows from (2) and Tschebycheff inequality that $M_n \to 1$ as $n \to \infty$, thus $C \leqq 1$. But if $C = 1$, we conclude from (2) and the central limit theorem that M_n does not tend to 0. Hence $C < 1$, and (1′) is proved.

By similar methods we can prove the following results: Let $2 < c < 4$. Put

$$M_n^{(c)} = m\left(x; \left|\sum_{k=1}^{n} f_k(x)\right| > n^{2/c}\right).$$

Then the necessary and sufficient condition for the convergence of $\sum_{k=1}^{\infty} M_n^{(c)}$

is that

$$\int_{-\infty}^{\infty} t\, dG(t) = 0, \qquad \int_{-\infty}^{\infty} |\, t\,|^{c}\, dG(t) < \infty.$$

If $c < 2$ then the necessary and sufficient condition for the convergence of $\sum_{n=1}^{\infty} M_n^{(c)}$ is that $\int_{-\infty}^{\infty} |\, t\,|^{c}\, dG(t) < \infty$.

Finally we can prove the following result: Assume that $\int_{-\infty}^{\infty} t\, dG(t) = 0$ and $\int_{-\infty}^{\infty} t^4\, dG(t) < \infty$. Then there exists a constant r so that

$$\sum_{n=1}^{\infty} m\left[x; \left|\sum_{k=1}^{n} f_k(x)\right| > n^{1/2} \cdot (\log n)^r\right] < \infty. \tag{17}$$

The case of the Rademacher functions shows that (17) can not be improved very much, in fact only the value of r could be improved

NOTES

This section is devoted to brief research and expository articles on methology and other short items.

BROWNIAN MOTION ON THE SURFACE OF THE 3-SPHERE

By Kôsaku Yosida

Mathematical Institute, Nagoya University

1. Introduction. Let S be a n-dimensional compact riemann space with the metric $ds^2 = g_{ij}(x)\,dx^i\,dx^j$ such that the totality G of the isometric transformations of S onto S constitutes a Lie group transitive on S. Consider a temporally homogeneous Markoff process by which $P(t, x, y)$, $t > 0$, is the transition probability that a point x is transferred to y after the elapse of t-unit time. We assume that $P(t, x, y)$ is a Baire function in (t, x, y) and continuous in t, then P satisfies Smoluchouski's equation

$$P(t + s, x, y) = \int_S P(t, x, z)P(s, z, y)\,dz \qquad (t, s > 0), \tag{1.1}$$

dz being the G-invariant measure $\sqrt{g(x)}dx^1\,dx^2 \cdots dx^n$, $g(x) = \det(g_{ij}(x))$, and

$$P(t, x, y) \geqq 0, \tag{1.2}$$

$$\int_S P(t, x, y)\,dy = 1. \tag{1.3}$$

The spatial homogeneity of the transition process may be defined by

$$P(t, Tx, Ty) = P(t, x, y) \qquad \text{for } T \in G. \tag{1.4}$$

The "continuity" of the transition process may be defined, following after A. Kolmogoroff and W. Feller,[1] as follows. Let $L_1(S)$ be the function space of integrable (with respect to dx) functions $f(x)$ on S, then, for those $f(x)$ which are dense in $L_1(S)$,

$$\begin{gathered}\frac{\partial f(t, x)}{\partial t} = A \cdot f(t, x), \qquad (t \geqq 0);\\ f(t, x) = \int_S f(y)P(t, y, x)\,dy, \qquad (t > 0), \qquad f(0, x) = f(x),\end{gathered} \tag{1.5}$$

where, with non-negative $b^{ij}(x)$

$$\begin{aligned}(Af)(x) = \frac{1}{\sqrt{g(x)}}\frac{\partial}{\partial x^i}\left(-\sqrt{g(x)}\,a^i(x)f(x)\right)\\ + \frac{1}{\sqrt{g(x)}}\frac{\partial^2}{\partial x^i \partial x^j}\left(\sqrt{g(x)}\,b^{ij}(x)f(x)\right).\end{aligned} \tag{1.6}$$

[1] A Kolmogoroff, "Zur Theorie der stetigen zufälligen Prozesse," *Math. Annalen*, Vol. 108 (1933); W Feller, "Zur Theorie der stochastischen Prozesse," *Math Annalen*, Vol 113 (1937)

The temporally and spatially homogeneous "continuous" Markoff process may, if it exists, be called a Brownian motion on the homogeneous space S. The purpose of the present note is to show that, under some derivability hypothesis concerning $a^i(x)$ and $b^{ij}(x)$, there exists one and (essentially) only one Brownian motion on the surface of the 3-sphere S^3.

I here express my hearty thanks to Dr. Kiyosi Itô who proposed to me the problem and discussed and much improved the manuscript.

2. The defining equation for the Brownian motion. The spatial homogeneity (1.4) is equivalent to the fact that A is commutative with every operator $\tilde{T}$ defined by

$$(\tilde{T}f)(x) = f(Tx), \qquad T \in G, \tag{2.1}$$

because we have

$$\int_S f(y)P(t, y, Tx)\, dy = \int_S f(Ty)P(t, Ty, Tx)\, dTy = \int_S f(Ty)P(t, y, x)\, dy.$$

The condition (2.1) is equivalent to

$$XA = AX \text{ for any infinitesimal operator } X = \xi^k(x)\frac{\partial}{\partial x^k} \tag{2.2}$$

induced on S by the infinitesimal operator of the Lie group G. Thus, assuming the derivability of $a^i(x)$ and $b^{ij}(x)$ of necessary orders, we obtain from (2.2) the conditions:

$$\xi^k(x)\frac{\partial}{\partial x^k}\left(\frac{1}{\sqrt{g(x)}}\frac{\partial G^i(x)}{\partial x^i}\right) = 0, \tag{2.3}$$

$$\left(G^i(x) = -\sqrt{g(x)}\, a^i(x) + \frac{\partial \sqrt{g(x)}\, b^{ij}(x)}{\partial x^j}\right),$$

$$\frac{1}{\sqrt{g(x)}}H^i(x)\frac{\partial \xi^k(x)}{\partial x^i} + b^{ij}(x)\frac{\partial^2 \xi^k(x)}{\partial x^i \partial x^j} = \xi^i(x)\frac{\partial}{\partial x^i}\left(\frac{1}{\sqrt{g(x)}}H^k(x)\right), \tag{2.4}$$

$$(H^i(x) = G^i(x) + \frac{\partial}{\partial x^j}(\sqrt{g(x)}\, b^{ij}(x)),$$

$$b^{ij}(x)\frac{\partial \xi^k(x)}{\partial x^j} + b^{kj}(x)\frac{\partial \xi^i(x)}{\partial x^j} = \xi^j(x)\frac{\partial b^{ik}(x)}{\partial x^j}. \tag{2.5}$$

Now for the surface of the 3-sphere S^3,

$$ds^2 = d\theta^2 + \sin^2\theta \cdot d\varphi^2, \qquad g(\theta, \varphi) = \sin^2\theta,$$

and the infinitesimal operators

$$X_x = \sin\varphi\frac{\partial}{\partial\theta} + \frac{\cos\theta\cos\varphi}{\sin\theta}\frac{\partial}{\partial\varphi},$$

$$X_y = \cos\varphi\frac{\partial}{\partial\theta} - \frac{\cos\theta\sin\varphi}{\sin\theta}\frac{\partial}{\partial\varphi},$$

$$X_z = \frac{\partial}{\partial \varphi}$$

respectively correspond to the rotations about the x-, y- and z-axis.

From (2.5) we see that, by taking $X = X_z$,

(2.6) $\qquad b^{ij}(\theta, \varphi)$ is independent of φ.

By taking $X = X_z$ in (2.4) we see that H^h is independent of φ. Hence, by (2.6),

(2.7) $\qquad a^i(\theta, \varphi)$ is independent of φ.

Thus, by taking $k = 1$, $X = X_x$ we obtain from (2.4),

$$\frac{1}{\sin\theta} H^2(\theta) \cos\varphi - b^{22}(\theta) \sin\varphi = \sin\varphi \frac{d}{d\theta}\left(\frac{1}{\sin\theta} H^1(\theta)\right)$$

and thus

(2.8) $$H^2(\theta) = 0, \qquad b^{22}(\theta) + \frac{d}{d\theta}\left(\frac{1}{\sin\theta} H^1(\theta)\right) = 0.$$

Hence, by taking $k = 2$, $X = X_x$ or $X = X_y$, we obtain from (2.4)

$$\frac{-H^1(\theta)\cos\varphi}{\sin^2\theta} + 2b^{11}(\theta)\frac{\cos\theta\cos\varphi}{\sin^2\theta} + 2b^{12}(\theta)\frac{\sin\varphi}{\sin\theta} - b^{22}(\theta)\frac{\cos\theta\cos\varphi}{\sin\theta} = 0,$$

$$\frac{H^1(\theta)\sin\varphi}{\sin^2\theta} - 2b^{11}(\theta)\frac{\cos\theta\sin\varphi}{\sin^2\theta} + 2b^{12}(\theta)\frac{\cos\varphi}{\sin^2\theta} + b^{22}(\theta)\frac{\cos\theta\sin\varphi}{\sin\theta} = 0.$$

From these two equations we obtain

(2.9) $$b^{12}(\theta) = 0, \qquad \frac{H^1(\theta)}{\sin^2\theta} - 2b^{11}(\theta)\frac{\cos\theta}{\sin^2\theta} + b^{22}(\theta)\frac{\cos\theta}{\sin\theta} = 0.$$

By taking $i = 2$, $k = 1$, $X = X_z$, we obtain from (2.5), (2.9)

$$b^{22}(\theta)\cos\varphi + b^{11}(\theta)\frac{d}{d\theta}\left(\frac{\cos\theta\cos\varphi}{\sin\theta}\right) = 0$$

and hence

(2.10) $$b^{22}(\theta) = \frac{b^{11}(\theta)}{\sin^2\theta}.$$

Similarly by taking $i = 1$, $k = 1$, $X = X_z$ we obtain from (2.5)

$$b^{12}(\theta)\cos\varphi + b^{12}(\theta)\cos\varphi = \sin\varphi \frac{db^{11}(\theta)}{d\theta}$$

and hence by (2.9), (2.10)

(2.11) $$b^{11}(\theta) = \text{constant } C, \qquad b^{22}(\theta) = \frac{C}{\sin^2\theta}.$$

Thus we obtain from (2.4)

$$H^1(\theta) = -a^1(\theta)\sin\theta + 2C\cos\theta, \qquad H^2(\theta) = -\sin\theta\cdot a^2(\theta)$$

and thus, by (2.8),

$$a^2(\theta) = 0. \tag{2.12}$$

Substituting (2.11) in (2.9) we obtain

$$a^1(\theta) = \frac{C\cos\theta}{\sin\theta}. \tag{2.13}$$

Therefore since $b^{11}(\theta)$ and $b^{22}(\theta)$ are non-negative, A is (essentially) equal to the Laplace operator

$$\Lambda = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}. \tag{2.14}$$

Thus we may obtain $P(t, x, y)$ by integrating the equation

$$\frac{\partial f(t;\theta,\varphi)}{\partial t} = \Lambda\cdot f(t;\theta,\varphi), \qquad (t \geqq 0), \tag{2.15}$$

and by putting

$$f(t;\theta,\varphi) = f(t, x) = \int_{S^3} f(y)P(t, y, x)\,dy. \tag{2.16}$$

3. Integration of the equation (2.15)–(2.16). Consider the Laplacian (real) spherical harmonics

$$Y_k^{(m)}(\theta,\varphi) = Y_k^{(m)}(x), \qquad (-k \leqq m \leqq k; k = 0, 1, \cdots). \tag{3.1}$$

They constitute an orthonormal function system complete for continuous functions on S^3, and we have

$$\Lambda\cdot Y_k^{(m)}(\theta,\varphi) = -k(k+1)Y_k^{(m)}(\theta,\varphi). \tag{3.2}$$

Since, as is well-known,

$$Y_k^{(m)}(T^{-1}x) = \sum_{n=-k}^{k} u_{nm}^{(k)}(T)Y_k^{(n)}(x) \tag{3.3}$$

by an irreducible orthogonal representation $(u_{nm}^{(k)}(T))$ of the rotation group G, we have

$$\max_x |\, Y_k^{(m)}(x)\,|^2 \leqq (2k+1)\min_x \sum_{n=-k}^{k} |\, Y_k^{(n)}(x)\,|^2, \tag{3.4}$$

by applying the Schwarz inequality and the transitivity of the group G on S^3. The right hand member satisfies, by the orthonormality

$$(2k+1)^2/(\text{area of } S^3). \tag{3.5}$$

Therefore the double series (for $t > 0$)

$$(3.6)\quad P(t;\theta,\varphi;\theta',\varphi') = \sum_{k=0}^{\infty}\sum_{m=-k}^{k} \exp\left(-k(k+1)t\right)Y_k^{(m)}(\theta,\varphi)Y_k^{(m)}(\theta',\varphi')$$

is absolutely and uniformly convergent on S^3. We will show that this P is the required (unique) Brownian motion on S^3.

The proof may be given in three steps. i) We see by (3.2) and (3.6), that $\int_{S_3} f(y)P(t, y, x)\,dx$ satisfies (2.15) if

$$f(x) \sim \sum_{k=0}^{\infty}\sum_{m=-k}^{k} d_k^{(m)}Y_k^{(m)}(x), \qquad \sum_{k=0}^{\infty}\sum_{m=-k}^{k} \exp\left(-k(k+1)t\right)k(k+1)\,d_k^{(m)}Y_k^{(m)}(x)$$

are both absolutely and uniformly convergent. By the completeness of $\{Y_k^{(m)}(x)\}$, such $f(x)$ are dense in $L_1(S)$.

ii) Because of (3.3) we see that (3.6) satisfies the spacial homogeneity (1.4).

iii) (1.3) is obvious by the orthonormality of $\{Y_k^{(m)}(x)\}$ and the constancy on S^3 of $Y_0^{(0)}(x)$. Next, for the solution $f(t, x)$ of (2.15)–(2.16), let $f(x) = f(0, x)$ be non-negative on S^3, then $g_\epsilon(t, x) = \exp(-\epsilon t)f(t, x)$, $(\epsilon > 0)$, satisfies

$$\frac{\partial g_\epsilon(t, x)}{\partial t} = \Lambda \cdot g_\epsilon(t, x) - \epsilon g_\epsilon(t, x), \qquad (t > 0),$$

$$g_\epsilon(0, x) = f(x) \geqq 0 \qquad (\text{on } S^3).$$

Thus $g_\epsilon(t, x) \geqq 0$ on S^3, since $g_\epsilon(t, x)$ cannot have a negative minimum on the product space $[t_1, t_2] \times S^3$, for any $t_2 > t_1 > 0$. For at such minimizing point we must have

$$\frac{\partial g_\epsilon}{\partial t} = 0, \qquad \frac{\partial g_\epsilon}{\partial \theta} = 0, \qquad \frac{\partial g_\epsilon}{\partial \varphi} = 0, \qquad \frac{\partial^2 g_\epsilon}{\partial \theta^2} \geqq 0, \qquad \frac{\partial^2 g_\epsilon}{\partial \varphi^2} \geqq 0.$$

Therefore, since $\epsilon > 0$, $t_2 > t_1 > 0$ were arbitrary, we conclude that $f(t, x) \geqq 0$ on S^3 for $t > 0$ if $f(x) = 0$ on S^3. This proves (1.2). The same argument simultaneously shows us that the solution P of (2.15)–(2 16) and (1.2)–(1.3) is unique.

ON THE STRONG STABILITY OF A SEQUENCE OF EVENTS

By Aryeh Dvoretzky

Hebrew University, Jerusalem, and Institute for Advanced Study

1. Summary. M. Loève [3] has found conditions under which a sequence of events which may be interdependent in an arbitrary manner is strongly stable. In this note it is established that considerably weaker conditions imply the strong stability.

2. Introduction. Let

$$(1)\qquad A_1, A_2, \cdots, A_n, \cdots$$

be a sequence of events, which may depend on each other in any way whatsoever, defined on the same set of trials.

Let R_n be the *repetition* function of (1), i.e. R_n is the number of those among the first n events: $A_1, A_2, \cdots, A_n$ which were realized, and put $f_n = R_n/n$. The random variable f_n is called the *frequency* function of (1).

Denoting by $E\{x\} = \bar{x}$ the expected value of x it is evident that

$$\bar{R}_n = E\{R_n\} = \sum_{\nu=1}^{n} \Pr(A_\nu), \qquad \bar{f}_n = E\{f_n\} = \frac{1}{n} E\{R_n\}.$$

Following Loève [3, p. 252] we say that (1) is *strongly stable* if the sequence $\varphi_n = f_n - \bar{f}_n$ $(n = 1, 2, \cdots)$ is strongly stable in the usual Kolmogoroff sense [1, p. 58], i.e. if

$$\lim_{n\to\infty} \Pr(\sup_{\nu > n} |\varphi_\nu| > \epsilon) = 0 \tag{2}$$

for every $\epsilon > 0$.

Putting[1]

$$\beta_n = \frac{1}{n} \sum_{i=1}^{n} \Pr(A_i), \qquad \gamma_n = \frac{2}{n(n-1)} \sum_{1 \le \mu < \nu \le n} \Pr(A_\mu A_\nu)$$

and introducing the abbreviation[2]

$$\delta_n = \gamma_n - \beta_n^2,$$

Loève's result [3, pp. 257–9] is the following:

If $n\delta_n$ is bounded then (1) *is strongly stable.*

This, even when specialized to sequences of independent events, includes the Bernoulli and Poisson cases.

Here the following stronger result will be established.

THEOREM. *If $\Sigma\, \delta_n/n$ is convergent then* (1) *is strongly stable.*

In particular, if for some $\epsilon > 0$ the sequence $n^{\epsilon}\delta_n$ is bounded then (1) is strongly stable.

3. A lemma. The new tool here used is the following simple result on series of positive terms.

LEMMA. *Let $a_n \ge 0$ for $n = 1, 2, \cdots$ and*

$$\sum_{n=1}^{\infty} \frac{a_n}{n} \tag{3}$$

be convergent. Then there exists a sequence n_i of integers satisfying

$$0 < n_{i+1} - n_i = o(n_i) \qquad (i \to \infty), \tag{4}$$

and such that the series $\Sigma_{i=1}^{\infty} a_{n_i}$ is convergent.

---

[1] $A_\mu A_\nu$ denotes the event: both A_μ and A_ν.

[2] Our β_n, γ_n and δ_n correspond to Loève's $p_1(n)$, $p_2(n)$ and d_n^2 respectively.

PROOF. Since (3) is convergent it is well known[3] that there exists a sequence of numbers $l_n (n = 1, 2, \cdots)$ satisfying

$$l_{n+1} \geq l_n, \qquad \lim_{n=\infty} l_n = \infty \tag{5}$$

having the property that

$$\sum_{n=1}^{\infty} l_n \frac{a_n}{n} < \infty. \tag{6}$$

We define inductively a sequence of integers $m(i)$ through

$$m(1) = 1, \qquad m(i+1) = m(i) + 1 + \left[\frac{m(i)}{l_{m(i)}}\right], \tag{7}$$

the square brackets denoting the integral part. Clearly

$$0 < m(i+1) - m(i) = o(m(i)). \tag{8}$$

Now for every i we choose n_i so that

$$m(i) \leq n_i < m(i+1) \quad \text{and} \quad a_{n_i} = \min_{m(i) \leq \nu < m(i+1)} a_\nu.$$

These n_i satisfy the requirements of the lemma.

Indeed, (4) holds in virtue of (8) while applying (5) and (7) we obtain

$$s_i = \sum_{\nu=m(i)}^{m(i+1)-1} l_\nu \frac{a_\nu}{\nu} \geq (m(i+1) - m(i)) l_{m(i)} \frac{a_{n_i}}{m(i+1)} \geq \frac{m(i)}{m(i+1)} a_{n_i}.$$

Since Σs_i converges by (6) it follows from the preceding inequality and (8) that $\Sigma a_{n_i} < \infty$ as required.

COROLLARY. *The conclusion of the lemma remains valid if the condition $a_n \geq 0$ is dropped provided (3) is absolutely convergent.*

4. Proof of the theorem. An easy calculation [3, p. 253] gives

$$\sigma_n^2 = E\{(f_n - \bar{f}_n)^2\} = \delta_n + \frac{\beta_n - \gamma_n}{n}.$$

Since both β_n and γ_n are between zero and one we have

$$-\frac{1}{n} < \sigma_n^2 - \delta_n < \frac{1}{n}.$$

Therefore it follows from the assumption of the theorem that $\Sigma (\sigma_n^2/n)$ is convergent. Hence by the lemma there exists a sequence of integers n_i satisfying (4) and such that $\Sigma \sigma_{n_i}^2$ converges.

---

[3] Take e.g. $l_n = (\Sigma_{\nu > n} \nu^{-1} a_\nu)^{-\frac{1}{2}}$ (cf [2, p. 299]).

Applying Tchebytcheff's inequality to $\varphi_{n_\nu} = f_{n_\nu} - \bar{f}_{n_\nu}$ and adding for $\nu \geq i$ we have for every $\epsilon > 0$

$$\Pr\left(\sup_{\nu \geq i} |\varphi_{n_\nu}| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{\nu=i}^{\infty} \sigma_{n_\nu}^2. \tag{9}$$

If $n_i \leq n < n_{i+1}$ then

$$|f_n - f_{n_i}| = \left|\frac{R_n}{n} - \frac{R_{n_i}}{n_i}\right| < \frac{n_{i+1} - n_i}{n_i}.$$

Denoting the last term of this inequality by ϵ_i and putting $\bar{\epsilon}_i = \max_{\nu \geq i} \epsilon_\nu$, we have from (9)

$$\Pr\left(\sup_{n \geq n_i} |\varphi_n| > \epsilon + 2\bar{\epsilon}_i\right) \leq \frac{1}{\epsilon^2} \sum_{\nu=i}^{\infty} \sigma_{n_\nu}^2.$$

As $\bar{\epsilon}_i \to 0$ and the right hand term is the remainder of a convergent series, (2) follows and the theorem is proved.

5. Remarks. 1. The lemma used here can also be applied to the study of the order of magnitude of φ_n in the almost certain sense.

2. If the terms of (3) are decreasing then the existence of a convergent subseries of Σa_n satisfying (4) implies $\Sigma_{i=1}^{\infty} a_{2^i} < \infty$. But this is equivalent to the convergence of the series with monotone terms (3) (cf. e.g. [2, p. 130]). Hence in this case the convergence of (3) is *necessary as well as sufficient* for the validity of the lemma. It may be possible to use this remark in order to establish in some special cases, where the interdependence of the variables decreases steadily in a suitable sense, necessary and sufficient conditions for strong stability.

3. The sequence of δ_n is of course, of very specialized structure. Thus, since the stability of (1) is equivalent [3, p. 255] to $\delta_n \to 0$ and is implied by strong stability, it follows that $\delta_n \to 0$ whenever $\Sigma (\delta_n/n)$ is convergent.

Added in proof: Since this paper was submitted I heard from Professor M. Loève that he has independently obtained the theorem of section 2 by another method.

REFERENCES

[1] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitrechnung*, Ergeb. d. Math Vol. 2, no. 3, Springer, Berlin, 1933.

[2] K. Knopp, *Theory and Applications of Infinite Series*, Blackie, London and Glasgow, 1928.

[3] M. Loève, "Étude asymptotique des sommes de variables aléatoires liées," *Jour. de Math. pures et appl.*, Vol. 24 (1945), pp. 249–318.

A NOTE ON WEIGHING DESIGN

By K. S. Banerjee

Pusa, Bihar, India

1. Efficiency of weighing designs given by a three-fourth replicate. In the June issue of the *Annals*, Kempthorne [1] approached the construction of the orthogonal matrix X through fractional replicates, the original treatment of which was given by Finney [2]. Reference has been made to the use of a three fourth replicate for weighing designs. Details for such designs have not been furnished as their efficiency is lower than for the designs given by the completely orthogonal matrix X In a three fourth replicate the treatment combinations have to be chosen in a particular manner for a comparatively easier analytical treatment both from the point of view of agrobiological experiments as well as weighing designs. The variance of each of the estimates in such a case will be $\sigma^2/2^{n-1}$. As a matter of fact, in a weighing design given by a fractional replicate of the type of $(2^\beta - 1)/2^\beta$, $(\beta = 1, 2, \cdots n)$, of 2^n experiments, the estimate of the variance of each object is independent of the fraction used and is equal to $\sigma^2/2^{n-1}$, the same as above.

2. Construction of a three fourth replicate. Kempthorne mentions that a factorial design of fraction $\frac{3}{4}$ could be taken to consist of a $\frac{1}{2}$ replicate on the identity $I = ABC$ and a quarter replicate based on the identity

$$I = A = BC = ABC.$$

If the half replicate based on the identity $I = ABC$ be taken to consist of all the treatments corresponding to the minus signs of the treatment contrast ABC [3], the additional quarter replicate can be chosen in two different ways. When however the treatments corresponding to the minus signs of both A and BC are kept, omitting the treatments corresponding to the plus signs of A and BC, the three fourth replicate so obtained will have certain advantages, which will not be available if the quarter replicate to be added is chosen to consist of the treatments corresponding to the plus signs of A and BC.

3. Behavior of the contrasts in a three fourth replicate and the efficiency of the weighing designs. In general, if there are n treatments giving rise to 2^n treatment combinations and if the defining contrasts be chosen as

$$I = ACD = BDE = ABCE,$$

it will be necessary to omit the treatment combinations corresponding to the plus signs of both ACD and BDE, which will be 2^{n-2} in number. In the three fourth replicate so obtained, 2^n treatment effects (inclusive of the mean) will divide themselves into sets of 4 treatment contrasts each. One of the sets will be I, ACD, BDE and $ABCE$ and any other set will be formed by multiplying any treatment contrast by the defining set namely, I, ACD, BDE and $ABCE$. Only three contrasts out of four in a set will be independent, so that only one of

the contrasts, preferably the one of the highest order interaction may be kept as an alias (in agrobiological experiments) of the remaining three and may therefore be omitted. Each of the four contrasts within a set will be orthogonal to each of the other contrasts in the remaining sets, but within a set the four contrasts will be non-orthogonal to one another. Though non-orthogonal, the normal equations will be of the systematic type[1] and the matrix $X'X$, taking any three contrasts out of each set of four, will take the following form:

$$(1) \qquad \begin{bmatrix} x & a & a & 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ a & x & a & 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ a & a & x & 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & x & a & a & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & a & x & a & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & a & a & x & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & x & a & a & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & a & x & a & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & a & a & x & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

where the order of the matrix $N = \frac{3}{4}2^n$ is of the form $3t (t = 2^{n-2})$ and $x = 3.2^{n-2}$, $a = -\frac{1}{2}2^n + \frac{1}{4}2^{n-2} = -2^{n-2}$. The value of the above determinant $= (x - a)^{2t}(x + 2a)^t$ and that of the determinant suppressing the first row and the first column $= (x - a)^{2t-1}(x + a)(x + 2a)^{t-1}$. $a'' = (x + a)/(x - a)(x + 2a) = 1/2^{n-1}$, substituting for x and a. The variance of each estimate will therefore be $\sigma^2/2^{n-1}$.

4. General case. When a fraction of the type $\alpha/2^\beta = (2^\beta - 1)/2^\beta$ is used, the treatment combinations corresponding to the plus signs of the β independent contrasts is omitted. Out of each set of 2^β treatment contrasts, only $\alpha = 2^\beta - 1$ will de independent and the matrix will then take a form like that of (1), where

$x = [(2^\beta - 1)2^n]/2^\beta = 2^{n-\beta}(2^\beta - 1)$ and

$a = -\frac{1}{2}2^n + [(2^{\beta-1} - 1)2^n]/2^\beta = -2^{n-\beta}$,

$a'' = [x + (\alpha - 2)a]/(x - a)\,[x + (\alpha - 1)a] = (2 \cdot 2^{n-\beta})/2^n 2^{n-\beta} = 1/2^{n-1}$.

The variance of each estimate $= \sigma^2/2^{n-1}$, the same as before. When a completely orthogonalised matrix of the order $(\alpha 2^n)/2^\beta = 2^{n-\beta}(2^\beta - 1)$ is available, the variance of an estimate will be $\sigma^2/2^{n-\beta}(2^\beta - 1)$. The ratio of the two variances $= 2^{n-1}/(2^n - 2^{n-\beta}) = 2^{\beta-1}/(2^\beta - 1)$, which shows how the efficiency of the weighing design decreases with the increasing value of the fraction. When $\beta = 1$, i.e. in a half replicate, the efficiency is 100 percent. The value of the fraction is never less than $\frac{1}{2}$.

[1] The analysis of the data available from agrobiological experiments will not be cumbersome to a prohibitive extent as in many other experiments where non-orthogonality creeps in The results of investigation in this direction have already been communicated for publication elsewhere.

5. Independence of the estimates given by L_N in a biased spring balance. Kempthorne mentions that although the optimum designs for the spring balance case suggested by Mood furnish somewhat smaller variance than what is given by fractional replicates, these designs have the disadvantage that the estimates are correlated, whereas the estimates furnished by fractional replicates are orthogonal. The designs furnished by fractional replicates take account of the bias and if the weighing operation corresponding to the bias is omitted (in case where the spring balance is free from bias), the resultant scheme will fail to give independent estimates and the variance factors will be of the same magnitude as in the optimum design L_N of Mood with the same number of weighings. Again, these optimum designs may also be made to furnish independent estimates when the designs are adjusted in the manner as suggested by Mood to suit a biased spring balance.

It is true that the design matrix L_3 given by

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

does not give independent estimates as such; but when it is assumed that the spring balance has a bias and the design matrix is modified as follows:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \tag{2}$$

the estimates except that for the bias will be orthogonal to one another and the variance of the estimated weights will necessarily be larger in value.

Before proving the general case, we notice that when -1 is substituted for 0 in (2) above, the resultant scheme will be an orthogonalised matrix. This is true not only in this particular instance but will hold good also in general. The constitution will be clear when the method of construction of L_N from H_{N+1} is recalled.

The distribution of ones in L_N gives a special type of symmetrical balanced incomplete block design, where $r = k = \frac{1}{2}(b + 1)$ and $\lambda = \frac{1}{4}(b + 1)$, while the distribution of zeros gives the complementary design for which $r_0 = r - 1$, $k_0 = k - 1$ and $\lambda_0 = \lambda - 1$. Therefore when a row of zeros and a column of ones (in that order) is added to L_N, the matrix $X'X$ of the resultant scheme takes the following form:

$$\begin{bmatrix} N+1 & r & r & r & \cdots & r \\ r & r & \lambda & \lambda & \cdots & \lambda \\ r & \lambda & r & \lambda & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r & \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}. \tag{3}$$

Making use of the identities well known in the theory of balanced incomplete block designs and remembering the relationships, $2\lambda = r = k = \frac{1}{2}(N+1)$,

(I) The value of the determinant of

$$X'X = (r-\lambda)^{N-1}[(N+1)\{r+\lambda(N-1)\} - r^2N] = (r-\lambda)^{N-1}[r+\lambda(N-1)],$$

(II) The value of the determinant suppressing the first row and the first column $= (r-\lambda)^{N-1}[r+\lambda(N-1)]$,

(III) The value suppressing the second row and the second column

$$= (r-\lambda)^{N-2}[(N+1)\{r+\lambda(N-2)\} - r^2(N-1)]$$
$$= (r-\lambda)^{N-2}[r+\lambda(N-1)],$$

(IV) The value suppressing the first row and the third column

$$= (r-\lambda)^{N-2}[r\{r+\lambda(N-2)\} - r\lambda(N-1)]$$
$$= r(r-\lambda)^{N-1},$$

(V) The value suppressing the second row and third column

$$= (r-\lambda)^{N-2}[\lambda(N+1) - r^2]$$
$$= 0.$$

Hence, the reciprocal matrix of $X'X$ will be given by

$$[X'X]^{-1} = \begin{bmatrix} 1 & -1/k & -1/k & \cdots & -1/k \\ -1/k & 2/k & 0 & \cdots & 0 \\ -1/k & 0 & 2/k & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1/k & 0 & 0 & \cdots & 2/k \end{bmatrix}. \tag{4}$$

Let Y' denote the column matrix of the results of the weighings, $y_0, y_1, \cdots, y_N$ and B' the column matrix of the estimates of the weights $b_0, b_1, \cdots b_N$. Then the estimates will be given by the equation

$$B' = [X'X]^{-1}X'Y'.$$

It is easy to see that all the rows except the first in $[X'X]^{-1}X'$ are orthogonal to one another. To explain this, let us take the design given by (2). Here

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Then $[X'X]^{-1}X'$ will be of the form

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1/k & +1/k & +1/k & -1/k \\ -1/k & +1/k & -1/k & +1/k \\ -1/k & -1/k & +1/k & +1/k \end{bmatrix}.$$

In all the rows excepting the first, for every 0 and $+1$ in X', there will respectively be a $-1/k$ and a $+1/k$ in $[X'X]^{-1}X'$. It has been mentioned before

that an orthogonal matrix is obtained when -1 is substituted for every 0 in X or X'. Hence, N rows (all except the first) of $[X'X]^{-1}X'$ will be orthogonal and these N rows will estimate the N weights in orthogonal linear combinations of $y_0, y_1 \cdots y_N$.

It has been mentioned before that the distribution of zeros in L_N gives the complementary design, for which $r_0 = r - 1$, $K_0 = k - 1$ and $\lambda_0 = \lambda - 1$. If to such a design, a row of ones and a column of ones (in that order) be added to suit the estimation of the weights in a biased spring balance, exactly a similar situation will be obtained and the estimates will be orthogonal. It can readily be seen that the design furnished by Yates to weigh seven light objects and a bias is an illustration of this kind. The scheme given by Yates is the complementary design of L_7 with an additional row and a column of ones added to L_7.

The sixteen combinations of ten objects, a, b, c, d, e, f, g, h, k, l include 1, which corresponds to weighing with empty pans or, in other words, which is devoted to estimating the bias. When 1 is omitted, $X'X$ will be of the form

$$\begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \lambda & \lambda & r & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda & \lambda & \lambda & & r \end{bmatrix},$$

where $r = 8$ and $\lambda = 4$. The above matrix $X'X$ is obviously of the same form as given by L_{15}.

By following exactly the same procedure as given above, it can easily be seen that when the weighing operation 1 is included in the weighing design, the solution of the normal equations will lead to independent estimates. The absence of each letter will be a 0 and the presence $a + 1$ in the design matrix and if -1 is substituted for every zero, the resultant matrix will be orthogonal. In some cases, however, the number of letters in all the combinations will not be the same, i.e. k will not be constant. In such a situation, k in (4) will take the value of r or of 2λ.

REFERENCES

[1] O. Kempthorne, "The factorial approach to the weighing problem," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 238–245.

[2] D. J. Finney, "The fractional replication of factorial arrangements," *Annals of Eugenics*, Vol. 12 (1945), pp. 291–301.

[3] F. Yates, *Tech. Commun. Bur. Soil Sci. Harpenden* no. 35 (1937), p. 11.

[4] Harold Hotelling, "Some improvements in weighing and other experimental techniques," *Annals of Math. Stat.*, Vol. 13 (1944), pp. 297–306

[5] K. Kishen, "On the design of experiments for weighing," *Annals of Math. Stat.*, (1945), pp. 294–301.

[6] A. M. Mood, "On Hotelling's weighing problem," *Annals of Math Stat.*, Vol. 17 (1946), pp. 432–446.

[7] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiment," *Biometrika*, Vol. 33 (1946), pp. 305–325.

CONTROL CHART FOR LARGEST AND SMALLEST VALUES

By John M. Howell

Los Angeles City College

1. Introduction. It may at times be desirable to use a control chart for largest and smallest values (L & S) in place of the conventional charts for averages and ranges ($\bar{X}$ & R). The chart for largest and smallest values has certain advantages: all information may be combined on one chart, computations are simple, and specifications may be placed on the chart. In this paper, constants for the use of this chart are developed and comparison is made with the average and range charts.

2. Constants for determining limits. Let L and S denote the largest and smallest values, respectively, in a sample of n pieces, and let $\bar{L}$ and $\bar{S}$ denote the averages of these values for k samples. Then $(\bar{L} + \bar{S})/2$ and $(\bar{L} - \bar{S})/d_2$ are unbiased estimates of the population mean and standard deviation, respectively, in the case of a random sample from a normal population. The value of the constant d_2 is given in [1] and repeated in table 1 for convenience. If we denote $(\bar{L} + \bar{S})/2$ by M and $(\bar{L} - \bar{S})$ by $\bar{R}$, control limits may be determined in terms of these statistics.

In conformance with usual control chart practice, we will set the upper control limit at $\bar{L} + 3\hat{\sigma}_L$ and the lower control limit at $\bar{S} - 3\hat{\sigma}_S$, where $\hat{\sigma}_L$ is an estimate of the standard deviation of the largest values in samples drawn from a normal population, and similarly for $\hat{\sigma}_S$. The results of Tippett [2] and Pearson [3] for $E(R)$ of samples from a normal population were used to determine expected values of L and S: $E(R) = d_2\sigma$. Here, R is the range of samples of size n: $R = L - S$. But since $E[(L + S)/2] = a$ for a symmetrical distribution, then $E(L) = a + d_2\sigma/2$ and $E(S) = a - d_2\sigma/2$, where a and σ are the mean and standard deviation of the normal population from which samples are drawn.

The probability element of the largest value [4] is given by:

$$n[F(L)]^{n-1}f(L)\,dL \quad \text{where} \quad f(x) = 1/\sqrt{2\pi}\,\sigma\, e^{-(x-a)^2/2\sigma^2} \quad \text{and} \quad F(x) = \int_{-\infty}^{x} f(y)\,dy.$$

Then $E(L^2) = n\int_{-\infty}^{\infty} L^2[F(L)]^{n-1}f(L)\,dL$. Integrals of this type, differing only by a constant factor have been evaluated by Hojo [5] and from his results d_4 was determined so that $\sigma_L = \sigma_S = d_4\sigma$. Values for d_4 for $n = 2, 5, 10$ are also given by Tippett [2]. "Three-sigma" control limits may then be given in the form: $M \pm A_3\bar{R}$, where $A_3 = 0.5 + 3d_4/d_2$. The expected value of the upper control limit will then be: $E(UCL) = a + A_4\sigma$, where $A_4 = (d_2/2) + 3d_4$. Values of these constants for various sample sizes are given in Table I.

In practice, it might be desired, in the case of control charts for individual measurements or for L and S, to have $E(UCL) = a + 3\sigma$, and the lower control limit symmetrically placed with respect to the central line. In this case, the formula for the limits would be: $M \pm 3\bar{R}/d_2$ or $M \pm \sqrt{n}A_2\bar{R}$, where $A_2 =$

$3/(d_2\sqrt{n})$ is given in [1]. Since the efficiency of M decreases rapidly with increasing sample size [6], it would probably be better to use $\bar{X}$ in place of M for determining the central line for a control chart when the sample size is greater than five. $\bar{X}$ is the "average of averages" as defined in [1].

The chart for largest and smallest values would then consist of a chart on which both the largest and smallest values are plotted, with the central line at M, and the limits as given above.

3. Comparison of charts for a particular case. A comparison of the L & S chart with the $\bar{X}$ chart for a particular case in which the sample size was three is given in Fig. 1. Measurements were the shear strength of spotweld coupons of

TABLE I

Constants for largest and smallest value chart

n	d_2	d_4	A_2	A_3	A_4	n
2	1.128	.825	1.880	2.72	3.03	2
3	1.693	.748	1.023	1.82	3.09	3
4	2.059	.709	.729	1.53	3.15	4
5	2.326	.670	.577	1.36	3.17	5
6	2.534	.648	.483	1.27	3.21	6
7	2.704	.627	.419	1.20	3.23	7
8	2.847	.614	.373	1.15	3.26	8
9	2.970	.600	.337	1.10	3.28	9
10	3.076	.588	.308	1.07	3.30	10

aluminum in pounds. Since the range chart had no points above the "three-sigma" control limit and showed no other peculiarities, it has been omitted.

4. General comparison of charts. We assume a mean of zero and a standard deviation of unity as a "given standard," and then compute the probabilities when the true values are a and σ. The probability of a point being inside of "3-sigma" control limits on the range chart under these conditions is: $P_1 = \Pr(R < d_2D_4/\sigma)$, where D_4 is given in [1]. The probabilities for the range used here were found from the Pearson-Hartley tables [3]. The usual normality assumptions are made.

The probability of a point being inside of "3-sigma" control limits on the average chart under the same conditions is:

$$P_2 = \int_{\sqrt{n}/\sigma\,((-3/\sqrt{n})-a)}^{\sqrt{n}/\sigma\,((3/\sqrt{n})-a)} \varphi(t)\,dt \quad \text{where} \quad \varphi(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}.$$

Since Daly [7] has shown that the average and range of samples from a normal

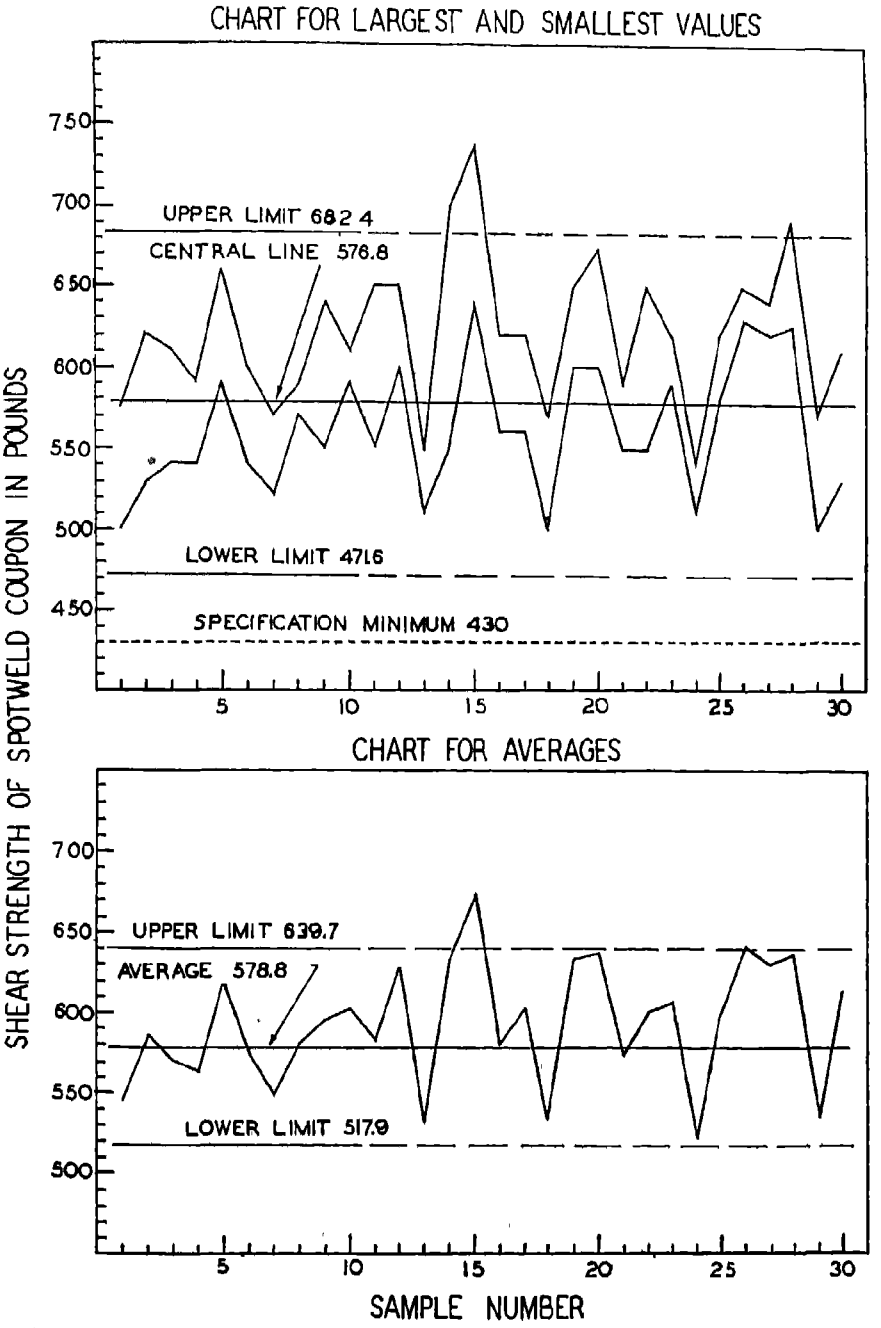
CHART FOR LARGEST AND SMALLEST VALUES
UPPER LIMIT 682.4
CENTRAL LINE 576.8
LOWER LIMIT 471.6
SPECIFICATION MINIMUM 430
CHART FOR AVERAGES
UPPER LIMIT 639.7
AVERAGE 578.8
LOWER LIMIT 517.9
SHEAR STRENGTH OF SPOTWELD COUPON IN POUNDS
SAMPLE NUMBER

FIG. 1

TABLE II

n	a	σ	P_1	P_2	P_1P_2	P_3	N_1	N_2
3	0	1.0	.994	.997	.991	.991	510	510
		1.2	.973	.988	.961	.963	116	122
		1.5	.901	.955	.860	.868	31	33
		2.0	.721	.866	.624	.645	10	11
3	0.5	1.0	.994	.983	.977	.980	198	228
		1.2	.973	.935	.935	.939	69	74
		1.5	.901	.917	.826	.834	25	27
		2.0	.721	.830	.598	.694	9	13
3	1.0	1.0	.994	.898	.893	.931	41	65
		1.2	.973	.855	.832	.860	25	31
		1.5	.901	.802	.723	.740	15	17
		2.0	.721	.746	.538	.550	8	8
3	2.0	1.0	994	.323	.321	.590	5	9
		1.2	.973	.352	.342	.510	5	7
		1.5	.901	.378	.341	.414	5	6
		2.0	.721	.408	.294	.321	4	5
5	0	1.0	.995	.997	.992	.992	570	570
		1.2	.969	.988	.957	.957	105	105
		1.5	.855	.955	.817	.878	23	36
		2.0	.588	.866	.509	.545	7	8
5	0.5	1.0	.995	.970	.965	.980	130	227
		1.2	.969	.942	913	.927	51	62
		1.5	.855	.891	.762	.791	17	20
		2.0	.588	.805	.473	.505	7	7
5	1.0	1.0	.995	.776	.722	.923	15	58
		1.2	.969	.736	.713	.828	14	25
		1.5	.855	.695	.594	.661	9	12
		2.0	.588	.648	.381	.426	5	6
5	2.0	1.0	.995	.071	.071	.512	2	7
		1.2	.969	.110	107	.402	3	6
		1.5	.855	.164	.140	.286	3	4
		2.0	.588	.230	.135	.185	3	3

population are independent, the probability that a sample is within control limits on both charts is the product of the probabilities: P_1P_2. Thus the probability that a sample be outside of control limits on either chart is $1 - P_1P_2$.

The probability of the largest and smallest values both lying in the interval from $-c$ to c is: $P_3 = \Pr(-c < S, L < c) = \left[\int_{(-c-a)/\sigma}^{(c-a)/\sigma} \varphi(t)\,dt\right]^n$. Values of this expression with lower limit $-\infty$ are given in table XXI of [8] for sample of sizes 3, 5, and 10. For the purpose of comparing the charts, we choose c so that the probabilities of Type 1 errors are equal, that is: $1 - P_1P_2 = 1 - P_3$ or $P_1P_2 = P_3$ when the mean is zero and the standard deviation unity. Substituting in this equation and solving, we find: $F(c) = 0.5 + 0.5\ (.9973P_1)^{1/n}$, where $F(x) = \int_{-\infty}^{x} \varphi(t)\,dt$. For $n = 3$, $c = 2.99$ and for $n = 5$, $c = 3.15$.

Comparing P_1P_2 with P_3 when the true values are a and σ will then show the relative power of the $\bar{X}$ & R charts and the L & S chart for detecting lack of control.

Finally the charts are compared by finding the number (N_1 for the $\bar{X}$ & R charts and N_3 for the L & S chart) of samples which will detect lack of control with a .99 probability under the conditions given above. This is done by finding the smallest integer which satisfies the following inequalities: $(P_1P_2)^{N_1} <$ .01 and $P_3^{N_3} <$ .01. As may be seen from table II, under most conditions, the L & S chart is nearly as good as the $\bar{X}$ & R charts for detecting lack of control.

REFERENCES

[1] American Standards Association, *Control Chart Method of Controlling Quality during Production*, Z1.3—1942.
[2] L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 18 (1925), pp. 364–387
[3] E. S. Pearson, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), pp. 301–308.
[4] S. S Wilks, *Mathematical Statistics*, Princeton University Press, 1943, p. 91.
[5] Hojo, "Distribution of median from a normal population," *Biometrika*, Vol. 23 (1931), p. 315
[6] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D Van Nostrand Co., 1931; p. 282.
[7] J. F. Daly, "On the use of the sample range in an analogue of Student's t-test," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71–74.
[8] Karl Pearson, *Tables for Statisticians and Biometricians*, Cambridge University Press, 1914.

SUFFICIENCY, TRUNCATION AND SELECTION[1]

By John W. Tukey

Princeton University

1. Summary. The fact that the mean and variance were sufficient statistics for a univariate normal distribution truncated at a fixed point was known to

[1] Prepared in connection with work sponsored by the Office of Naval Research.

Fisher by 1931 [2]. Hotelling [3] has recently observed the corresponding fact for the truncated multivariate normal distribution.

It is the aim of this note to point out that these are special cases of a general result, namely: *If a family of distributions admits a set of sufficient statistics, then the family obtained by truncation to a fixed set, or by fixed selection, also admits the SAME set of sufficient statistics.*

2. Representation. The basic formal results about sets of sufficient statistics are due to Fisher [1], whose arguments, with obvious modifications, establish that families of distributions satisfying the usual conditions have sufficient statistics. The converse was established by Koopman [4] for a reasonably wide class of families.

The usual condition can be easily handled and given wide application by representing the family of distributions in a form suggested to the author by Rubin, and ascribed by him to Cramér, namely:

$$dF(x \mid \theta) = c(\theta)f(x \mid \theta)\, d\mu(x),$$

where x is a possibly multidimensional chance quantity (i.e. random variable), θ is a possibly multidimensional parameter, $c(\theta)$ is a positive real function of θ which serves to normalize the distribution, $f(x \mid \theta)$—the relative probability density—is a non-negative real function of x and θ, and $\mu(x)$ is a positive measure function. In this representation the natural and sufficient condition that $\{h_i(x)\}$ are a set of sufficient statistics for θ is the existence of functions $a_i(\theta)$ such that (cf. Koopman [4])

$$\frac{\partial \log f(x \mid \theta)}{\partial \theta} = \sum_i a_i(\theta)h_i(x). \tag{1}$$

When θ is a vector, the derivative is to be interpreted as the gradient (a vector) and the $a_i(\theta)$ are to be vector-valued functions of θ. We notice that this condition concerns only the *relative* density function.

3. Proof of result. Suppose the family $F(x \mid \theta)$ is truncated onto a Borel set E, this means that

$$\Pr\{x \text{ in } E_1 \mid F(x \mid \theta) \text{ truncated to } E\} = \frac{\Pr\{x \text{ in } E \cap E_1 \mid F(x \mid \theta)\}}{\Pr\{x \text{ in } E \mid F(x \mid \theta)\}}.$$

If $\phi_E(x)$ is the characteristic function of E, which is $=1$ for x in ϵ and $=0$ otherwise, and if

$$k(\theta) = \Pr\{x \text{ in } E \mid F(x \mid \theta)\} = \int_E dF(x \mid \theta),$$

then the probability element of $F(x \mid \theta)$ truncated to E is

$$c(\theta)/k(\theta)f(x \mid \theta)\phi_E(x)\, d\mu(x) = c'(\theta)f(x \mid \theta)\, d\nu(x),$$

where $c'(\theta) = c(\theta)/k(\theta)$ and $dv(x) = \phi_x(x)\, d\mu(x)$. Truncation has not changed the relative density function, and the result follows from the form of (1).

Next suppose that, instead of accepting values with probability one in E and with probability zero outside E, we select according to a fixed Borel function $\phi(x)$, the chance of accepting a value x being $\phi(x)$. The new family of distributions has the same sufficient statistics for the same reason.

REFERENCES

[1] R. A. Fisher, "Theory of statistical estimation," *Camb Phil Soc Proc*, Vol 22 (1923-25), pp. 700–725

[2] R. A. Fisher, "The sampling error of estimated deviates together with other illustrations of the properties and applications of the integrals and derivatives of the normal error function," *Brit. Assn. Adv. Sci. Mathematical Tables*, Vol. 1, xxvi–xxxv.

[3] H. Hotelling, "Abstracts of Madison Meeting," *Annals of Math. Stat*, Vol. 19 (1948).

[4] B. O. Koopman, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39, pp. 399–409.

ON A PROBABILITY DISTRIBUTION

By Max A. Woodbury

University of Michigan

1. Introduction. The problem treated is that of generalizing the Bernouilli distribution to the case where the probability of success is not constant from trial to trial but depends on the number of previous successes. The case where the probability of an event depends on the number of trials is easily handled and is not the case treated here. Several special cases of such a distribution have been worked out at one time or another. (E.g C. C. Craig found the solution for one such special case and thus called the author's attention to the problem.)

The solution involves the Newton divided difference expansion of powers in a form which can be utilized for computation if the number of trials is not too large. In the case where the probabilities on a single trial are small an approximation, (similar to that of the Poisson distribution to the Bernouilli distribution) can be found.

Applications can obviously be made to urn schema in which black balls are replaced, but white balls are removed. Similarly, applications can be made to the distribution of the number of plants in a given area.

2. Solution of the problem. Specifically the problem is as follows: "What is the probability that in n trials of an event it will occur x times presuming that the probability of the event on a given trial depends only on the number of previous successes?" Denote by $P(n, x)$ the probability of x successes in n trials and by p_x the probability of the event after x previous successes. As

conventional denote $q_x = 1 - p_x$ and one can formulate the following equation of partial differences:

$$P(n+1, x+1) = p_x P(n, x) + q_{x+1} P(n, x+1). \tag{1}$$

This equation is an obvious consequence of the statement that $x+1$ successes in $n+1$ trials can only occur if there are x successes in n trials and a success on the $n+1$st or $x+1$ successes in n trials and failure on the $n+1$st. The boundary conditions appropriate are:

$$P(n, x) = 0 \text{ for } x < 0, \text{ or } x > n \text{ and } P(0, 0) = 1. \tag{2}$$

It is convenient and appropriate to generalize (1) while retaining the boundary conditions (2). The equation (1) will be obtained from the following equation by setting $q = 1$:

$$P(n+1, x+1) = (q - q_x)P(n, x) + q_{x+1} P(n, x+1). \tag{3}$$

It will be noted for further reference at this point that:

$$P(n, 0) = q_0^n \tag{4}$$

and:

$$P(n, n) = (q - q_0)(q - q_1) \cdots (q - q_{n-1}). \tag{5}$$

This last suggests a change of variable of the form:

$$P(n, x) = F(n, x)(q - q_0)(q - q_1) \cdots (q - q_x). \tag{6}$$

Upon substituting this expression in (3) one obtains a somewhat simpler equation with the same boundary conditions as (2).

$$F(n+1, x+1) = F(n, x) + q_{x+1} F(n, x+1). \tag{7}$$

Using the generating function:

$$G(x, \xi) = \sum_{n=x}^{\infty} F(n, x)\xi^n \tag{8}$$

one may obtain from (7), using the boundary conditions (2) the following ordinary linear difference equation:

$$G(x+1, \xi) = \xi[G(x, \xi) + q_{x+1} G(x+1, \xi)]. \tag{9}$$

From (4) it is easily seen that:

$$G(0, \xi) = 1/[1 - q_0\xi], \tag{10}$$

and hence that the solution of (9) is:

$$G(x, \xi) = \xi^x/[(1 - q_0\xi)(1 - q_1\xi) \cdots (1 - q_x\xi)]. \tag{11}$$

This may be expanded in partial fractions and the result written:

$$G(x, \xi) = \xi^x \sum_{i=0}^{x} q_i^x/[(q_i - q_0) \cdots (q_i - q_{i-1})(q_i - q_{i+1}) \cdots (q_i - q_x)(1 - q_i\xi)]. \tag{12}$$

By means of the relation in (8) one deduces readily that:

$$(13)\quad F(n, x) = \sum_{i=0}^{x} q_i^n / [(q_i - q_0) \cdots (q_i - q_{i-1})(q_i - q_{i+1}) \cdots (q_i - q_x)].$$

Jordan [1, p. 19, eq. (1)] shows this to be the xth Newton divided difference of q^n where the expansion is in terms of $(q - q_0) \cdots (q - q_x)$, for $x = 0, 1, \cdots, n$. The solution for (3) can now be written as:

$$(14)\quad P(n, x) = (q - q_0) \cdots (q - q_{x-1}) F_n(x)$$

from which follows:

$$(15)\quad \sum_{x=0}^{n} P(n, x) = q^n.$$

As remarked before, by setting $q = 1$ one obtains the solution of (1) subject to the boundary conditions (2).

It is clear that when all the q_i are equal that the Bernoulli distribution should come out as a special case. Since in this case the divided difference becomes the corresponding derivative divided by the appropriate factorial, one obtains:

$$(16)\quad P(n, x) = \frac{(1 - q_0)^x}{x!} \frac{d^x q^n}{dq^x}\bigg|_{q=q_0}.$$

Upon reduction this yields the usual formula, but not in the usual way.

By choosing $p_x = \lambda_x/n$ and allowing n to increase without limit one obtains an analogue of the Poisson distribution, viz:

$$(17)\quad P(x) = (-\lambda_0) \cdots (-\lambda_x) \sum_{i=0}^{x} e^{-\lambda_i} / [(\lambda_0 - \lambda_i) \cdots (\lambda_{i-1} - \lambda_i)(\lambda_{i+1} - \lambda_i) \cdots (\lambda_x - \lambda_i)]$$

which corresponds to the expansion of $e^{-\lambda}$ about $\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_x, \cdots$ when $\lambda = 0$.

REFERENCE

[1] Charles Jordan, *Calculus of Finite Differences*, Chelsea Publishing Co., New York, 2nd ed., 1947.

A GRAPHICAL DETERMINATION OF SAMPLE SIZE FOR WILKS' TOLERANCE LIMITS

By Z. W. Birnbaum and H. S. Zuckerman

University of Washington

1. Summary. To determine the smallest sample size for which the minimum and the maximum of a sample are the $100\beta\%$ distribution-free tolerance limits at the probability level ϵ, one has to solve the equation

$$(1)\quad N\beta^{N-1} - (N - 1)\beta^N = 1 - \epsilon$$

given by S. S. Wilks [1]. A direct numerical solution of (1) by trial requires rather laborious tabulations. An approximate formula for the solution has been indicated by H. Scheffé and J. W. Tukey [2], however an analytic proof for this approximation does not seem to be available. The present note describes a graph which makes it possible to solve (1) with sufficient accuracy for all practically useful values of β and ϵ.

2. Construction of the graph. Substituting in (1)

$$N = \frac{\beta}{1-\beta}\, x$$

we obtain

$$1 + x = (1 - \epsilon)\beta^{-\frac{\beta}{1-\beta}} x$$

and

$$\log (1 + x) = -\log \frac{1}{1-\epsilon} + \left(\frac{\beta}{1-\beta} \log \frac{1}{\beta}\right) x. \tag{2}$$

To solve (2) graphically, one has to find the intersection of the curve

$$y = \log (1 + x) \tag{3}$$

with the line

$$y = -\log \frac{1}{1-\epsilon} + \left(\frac{\beta}{1-\beta} \log \frac{1}{\beta}\right) x.$$

To prepare a graph on which this can be done, one first plots (3) once for al (Figure 1, Curve C). Then one marks the points $-\log \frac{1}{1-\epsilon}$ on the y-axis and labels them with the values of ϵ (Figure 1, Scale I); chooses a constant $r > 0$ and marks the points $r \log \frac{1}{1-\epsilon}$ on the x-axis (Figure 1, Scale II); chooses a constant $k > 0$, marks the points $kr \frac{\beta}{1-\beta} \log \frac{1}{\beta}$ on the x-axis, draws vertical lines through each of these points, and labels them with the values of β (Figure 1, Scale III); draws the line $x = k$ (Figure 1, line L); marks the uniform Scale IV on the x-axis.

The graph reproduced here has been prepared with $r = 4$, $k = 5$. It can easily be verified that the instructions on the graph lead to solutions x of (2) and $N = x \frac{\beta}{1-\beta}$ of (1).

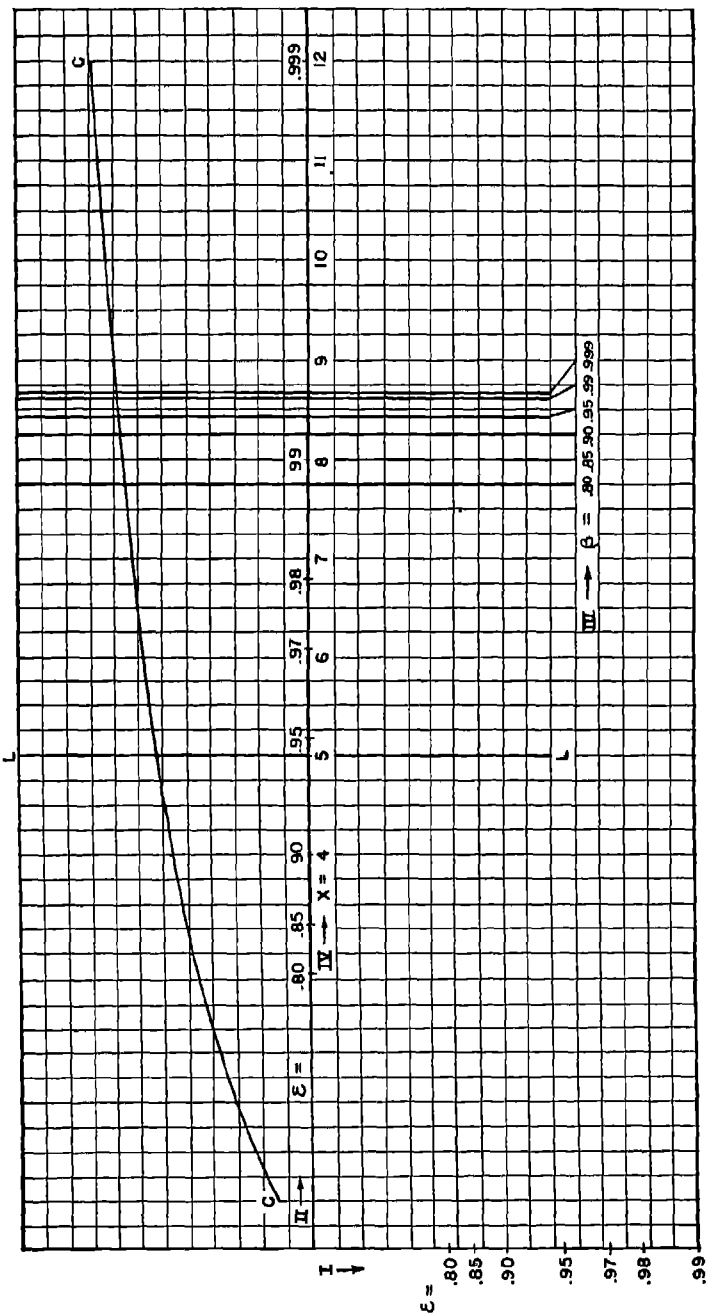

FIG. 1

To find an approximate solution of the equation $Np^{N-1} - (N-1)p^N = 1 - \epsilon$

1) connect ϵ on Scale I and ϵ on Scale II with a straight line; this line cuts vertical line marked β on Scale III at point P,
2) locate on line L the point with the ordinate of P; call this point Q,
3) connect ϵ on Scale I with Q; the connecting line cuts curve C at a point which has abscissa x on Scale IV; read off x,
4) compute $N = x\dfrac{\beta}{1-\beta}$

3. Improvement by iterations. The graphical solution, usually accurate to two significant digits, may be improved easily by iterations. Replacing (2) by the equation

$$x = \left[\log (1 + x) + \log \frac{1}{1 - \epsilon}\right]\left(\frac{\beta}{1 - \beta} \log \frac{1}{\beta}\right)^{-1} = f(x) \tag{4}$$

one obtains iterations $x_{i+1} = f(x_i)$ which, for $.80 \leq \epsilon \leq .999$ and $.80 \leq \beta \leq 999$ converge rapidly to the solution of (2).

EXAMPLE. For $\epsilon = .99$, $\beta = .999$, one finds graphically $x_1 = 6.6$, and from (4) the iteration formula $x_{i+1} = \dfrac{\log (1 + x_i) + 2}{.4337}$ which yields the values $x_2 = 6\,642$, $x_3 = 6.648$, $x_4 = 6.649$, $x_5 = 6.649$. Rounding up we obtain the sample size $N = 6.649 \cdot 999 = 6643$.

For ϵ and β between .80 and .999 all iterations obtained from (4) are on the same side of the exact solution and converge to it monotonically. Thus, in our example, from $x_1 < x_2$ we conclude that x_1 as well as all further iterations are smaller than the exact solution.

REFERENCES

[1] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943, p. 94.

[2] H. SCHEFFÉ AND J. W. TUKEY, "A formula for sample sizes for population tolerance limits," *Annals of Math. Stat.*, Vol. 15 (1944), p. 217.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the New York meeting of the Institute on April 8-9, 1949)

1. **Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix.** JACK SHERMAN and WINIFRED J. MORRISON, The Texas Company Research Laboratories, Beacon, New York.

If one element, a_{RS}, in a square matrix A is changed by an amount Δa_{RS}, all the elements b_{ij} in the inverse matrix B are generally changed. A simple equation has been derived by means of which the elements b_{ij} in the resulting inverse matrix B' can be computed directly in terms of Δa_{RS} and the elements of B. The equation is

$$b'_{ij} = b_{ij} - \frac{b_{Sj} b_{iR} \Delta a_{RS}}{1 + b_{SR} \Delta a_{RS}}$$

It follows that any given square matrix can be transformed into a singular matrix by increasing any one element in the transposed inverse matrix.

2. **The Distribution of the Number of Exceedances.** E. J. GUMBEL, New York and H. VON SCHELLING, Naval Research Laboratory, New London, Conn.

The probability for the mth observation in a sample of size n taken from a population with an unknown distribution of a continuous variate to be exceeded x times in N future trials is studied. The averages, moments, and the cumulative probability of the number of exceedances are calculated with the help of the hypergeometric series. The tolerance limits constructed by Wilks are special cases of the cumulative probability. The mean number of exceedances is the same as in Bernoulli's distribution. In some cases there are two modes, namely $m - 1$ and $m - 2$. If $n = N$, the most probable number of exceedances over the mth largest value is either m, or $m - 1$, and the median number of exceedances is equal to $m - 1$. In 50% of all cases, the largest (smallest) of n past observations will not (always) be exceeded in n future observations. If n and N are both large and equal, the distribution of the number of exceedances over the median is normal whereas the distribution of the extremes, similar to Poisson's distribution, has a mean m, and a variance $2m$. The variance of the number of exceedances is largest for the median, and smallest for the extremes of the previous sample. These distribution-free methods may be applied to meteorological phenomena, such as floods, droughts, extreme temperatures (the killing frost), largest precipitations, etc., and permit the forecasting of the number of cases surpassing a given severity.

3. **Note on the Power Function of a Quality Control Chart.** LEO A. AROIAN, Hunter College, New York.

The power function of a quality control chart is given for a sequence of N sample points in terms of α and γ, the probability of a Type I error and the power function respectively for a single sample point. Two different models are considered and the generalization to two quality control charts is indicated.

4. **Tests Between Two Means or Regression Coefficients When Observations are of Unequal Precision.** UTTAM CHAND, University of North Carolina, Chapel Hill.

Relative merits of different tests available for testing two means or two regression coefficients in relation to asymmetric and symmetric aspects of Student's hypothesis in case of unequal population variances have been reconsidered In this connection the distribu-

tion of a certain quantity t_k where k is some inexact value of the unknown ratio of variances has been obtained. The hypothesis of the equality of two linear regression functions in case of unequal residual variances has also been considered.

5. Functional Expansions. EUGENE W. PIKE, Boston, Massachusetts.

This paper calls attention to a new type of estimation problem, arising both in the interpretation of experimental data from complex experiments, and in the design of analogue computers for functions of several independent variables.

It has long been known, though not widely recognized, that the partial sums of rows and columns arising in the bivariate analysis of variance represent the least squares fit of a functional form $[f(x) + g(y)]$ to a tabular function $F(x, y)$ of two independent variables, for example More recently, several people have realized gradually that independent causes may combine in much more complicated ways to produce a common effect, and that correspondingly more complicated functional combinations, such as $[f(x) + g(y) + h(x)\cdot k(y)]$, can be fitted by least squares to tabular functions of x and y.

Examples of such expansions, as applied both to the design of computers and to the analysis of experimental data, will be given.

This presentation is based on work supported by the Air Materiel Command, USAF.

6. The Geometric Range for Distributions of Cauchy's Type. E. J. GUMBEL, New York City, and R. D. KEENEY, Metropolitan Life Insurance Comany, New York City.

From each of N samples of large size n the largest and the smallest values $X_{n,\nu}$ and $X_{1,\nu}$ $(\nu = 1, 2, \cdots N)$ are taken, where each X is measured from the central value of Nn observations. The sample size must be so large that the probability of any extreme $X_{n,\nu}$ and $-X_{1,\nu}$ being negative may be neglected. The distribution of the geometric means ρ of the N pairs of extremes henceforth called geometric ranges, is derived under the assumption that the initial distribution is symmetric, unlimited and of the Cauchy type which implies that the moments of an order equal to, or larger than $k (k > 0)$ diverge. Let u be the expected largest value. Then the probability density of $\xi_k = 2u^k\rho^{-k}$ obtained from a theorem of Elfving (*Biometrika*, Vol 35) is $\xi_k K_0(\xi_k)$ where K_0 is a Bessel function. This permits calculation of all moments of ξ_k. Methods are given for estimating the parameters u and k The distribution of the geometric ranges ρ is again a Bessel function. A probability paper is constructed for testing the hypothesis that the initial distribution is of Cauchy's type. A strict parallelism is established between the asymptotic distributions of the range for the exponential type, and of the geometric range for Cauchy's type. This provides a criterion to which of the two types the initial distribution belongs.

7. On Sums of Random Integers Reduced Modulo m. A. DVORETZKY, Institute for Advanced Study, Princeton and J. WOLFOWITZ, Columbia University, New York City.

Let X_n, $(n = 1, 2, \cdots)$ be an infinite sequence of independent, integral-valued, chance variables, and let m be any fixed integer greater than 1. Put $S_n = \sum_{\nu=1}^{n} X_\nu$ and denote S_n reduced mod. m by Y_n ; i.e., Y_n is a random variable which assumes only the values $j = 1, 2, \cdots, m$ with respective probabilities $P_n(j) = \text{Prob}\{S_n \equiv j \pmod m\}$. Necessary and sufficient conditions are obtained for Y_n to be equidistributed in the limit, i.e , for $\lim_{n=\infty} P_n(j) = \frac{1}{m}$. $(j = 1, 2, \cdots, m.)$ Some easily applicable sufficient conditions are deduced and the cases $m = 2, 3, 4$ are studied in detail. The rapidity with which $P_n(j) \to \frac{1}{m}$ is also studied

8. **The Corpuscle Problem: Estimating the Surface-Volume Ratio of a Corpuscle of Arbitrary Shape.** JEROME CORNFIELD, National Institutes of Health and HAROLD W. CHALKLEY, National Cancer Institute, Bethesda, Md.

Consider a space containing F, a closed figure of arbitrary shape, volume V and surface area S. Let a line segment of length r be thrown in the space in such a fashion that we have uniform distribution of the probabilities that the end point P occupies any position in the space and that the other end point P' occupies any position on the surface of a sphere of radius r with center at P. Count the number of end points falling in F (0, 1 or 2 for a single throw), call it the number of hits, and denote it by h. Count the number of times the line intersects the surface (0, 1 or 2 times for a single throw for a non-reentrant figure, possibly more for a re-entrant one), call it the number of cuts and denote it by c. Then, it is proved that $rE(h)/E(c) = 4V/S$ This result is intended to provide a theoretical basis for estimating the surface-volume ratio of physical objects of any shape.

9. **Generalized Hit Probabilities with a Gaussian Target.** D. A. S. FRASER, Princeton University.

In the *Supplement to the Journal of the Royal Statistical Society*, Vol. 8 (1946), L. B C. Cunningham and W. R B. Hynd proposed a problem and gave an approximate solution covering a partial range of parameter values: to find the probability that a moving target will survive a burst of "n" rounds from a rapid-firing gun, account being taken of correlation between the different points of aim.

Generalizing from the case of a two dimensional target to "k" dimensions, this paper gives the probability for 0, 1, 2, $\cdots$ n hits, under the following assumptions the "n" points of aim have a Multivariate Gaussian Distribution, the dispersion error has a Gaussian Distribution, and the target is a Gaussian Diffuse Target, that is, the probability of a hit on a particular round as a function of the coordinates of the shell has the form of "a constant times a Gaussian probability density function."

Limiting distributions are obtained as $n \to \infty$, subject to a variety of limiting conditions.

Numerical values for the probability of at least one hit are plotted when $n = 5$, for a range of values, relative to the target size, of dispersion and aiming errors

10. **A New Continuous Sampling Inspection Plan Based on an Analysis of Costs.** F. E. SATTERTHWAITE, General Electric Company, Bridgeport, Connecticut.

Inspection, like all other industrial operations, must be run to produce the most return for the lowest cost. The costs include overhead and running inspection costs; complaint costs; rework and scrap costs; and the costs of unnecessary process rejections Also one must consider the frequencies of occurrence of these costs. These include the process average percent defective; the probability of occurrence of a complaint, and the frequency of occurrence of quality deteriorations.

For continuous inspection, the percentage of the product to be inspected has a very simple formula: $P = \sqrt{SC/HM}$, where S is the sensitivity of the sampling plan used, C is the complaint cost, H is the effective inspection cost, and $1/M$ is the quality deterioration rate.

It was also necessary to develop a new continuous sampling inspection plan which would be efficient over the *entire* range of continuous sampling applications. The plan presented is a sequential plan which, with suitable attention to details, is easily applied on the shop floor. The Dodge Plan is a special case and is efficient only in a small percentage of applications.

11. **On the Levels of Significance of the F and Beta Distributions.** LEO A. AROIAN, Hunter College, New York.

Two formulas are given for the determinations of the levels of significance of the F and Beta distributions In the case of the F distribution a previous set of formulas (*Biometrika*, Vol. 34, pp 359-360) is modified to give 3 significant figure accuracy, n_1 , $n_2 \geqq 24$ The set for the Beta distribution is of Cornish-Fisher type, $p, q \geqq 6$. The advantage of these over Paulson's F formula and Carter's z formula are the avoidance of the solution of a quadratic in the case of Paulson's formula, and the avoidance of the exponential tables in the case of Carter's z formula A short numerical table compares the three methods for selected values of n_1 and n_2 .

12. **Certain Statistics for Samples of 3 From a Rectangular Population,** JULIUS LIEBLEIN, National Bureau of Standards.

A continuation of a study presented at the Madison meeting of the Institute of Mathematical Statistics last September (For abstract see *Annals of Math. Stat.*, December 1948, p 595.) The previous paper derived properties of the statistics

$$y_1 = \frac{x' - x''}{x_3 - x_1}, \qquad y_2 = \frac{x' - x''}{2}, \qquad y_3 = \frac{x' + x''}{2},$$

where x_1 , x_2 , x_3 are the observations, ordered by increasing size, in an independent random sample of three observations from a *normal* population, and x' and x'', $x' \geq x''$, are the two *closest* of the three. In the present paper distributions (joint as well as simple) are obtained for the above three statistics and also for x''', the remaining observation not included in the closest pair, for samples of 3 from a *rectangular* population, and a theorem is proved concerning the distribution of y_1 for a wide class of continuous populations.

13. **The Choice of Lot Inspection Plans of the Basis of Cost.** F. E. SATTERTHWAITE, and BURTON GRAD, General Electric Company, Bridgeport, Connecticut.

An extension of the first paper to single sampling inspection plans. The important concepts involved are the break-even quality level, the operating ratio, and the weighted prior odds that a lot is a good lot Charts are being prepared which can be entered with simple functions of the costs and which give directly the sample size and acceptance number for the most efficient single sampling inspection plan.

It appears promising that the method can be extended to double and sequential sampling plans This is imperative because of the large portion of the time that "no-inspection" is the most efficient single sampling plan.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Enrique Loizelier Blanco, Professor of Statistics in the University of Madrid, has just finished the first year of experimentation in Quality Control Methods in different plants The interest for these new statistical applications started in Spain during 1946 and have increased rapidly since then, especially this year after consecutive bimonthly intensive courses which Professor Blanco has been teaching.

Mr. Osmer Carpenter, formerly an Instructor in the Department of Statistics and Mathematics at Iowa State College is now doing statistical work for Carbide and Carbon Chemical Corp., Oak Ridge, Tennessee.

Dr. K. L. Chung, formerly of Princeton University, has been appointed to an assistant professorship at Cornell University.

Dr. Clyde H. Coombs, Associate Professor of Psychology and Chief of Research Division, Bureau of Psychological Services at the University of Michigan, is on leave of absence for the academic year to work at Harvard University on problems of scaling.

Dr. Meyer A. Girshick, formerly with the Douglas Aircraft Co., Santa Monica, California, has accepted a professorship in the Department of Statistics, Stanford University, Stanford, California.

Dr. M. J. Gottlieb, who has been with the Institute for Advanced Study at Princeton, has been appointed to an assistant professorship at the Newark College of Rutgers University.

Associate Professor E. H. C. Hildebrandt of Northwestern University has been elected President of the National Council of Teachers of Mathematics. He is also National Secretary-Treasurer of Pi Mu Epsilon and Secretary of the Mathematics Section of the Central Association of Science and Mathematics Teachers

Dr. C. A. Hollingsworth, formerly with the Acetate Section of the DuPont Company, is now an instructor in the Department of Chemistry, University of Pittsburgh.

Professor William G. Madow, who has been with the Institute of Statistics at the University of North Carolina, has been appointed Professor of Statistics at the University of Illinois.

Dr. Zenon Szatrowski, formerly teaching in the Economics Department of Northwestern University, has accepted an associate professorship in the Department of Economics, University of Oregon, Eugene, Oregon.

Mr. Eric Weyl has resigned his position as staff engineer in the Chicopee Manufacturing Corporation and is now conducting his own business as a textile engineering consultant in Manchester, New Hampshire.

New Members

The following persons have been elected to membership in the Institute (December 1, 1948 to February 28, 1949).

Abruzzi, Adam, M.S (Columbia Univ.) Student in engineering at Columbia University, *22 W. 107th Street, Shanks Village, New York.*

Agarwal, Satya P., M.A. (Agra Univ., India) Student at University of California, *International House, Berkeley 4, California.*

Anderson, Robert W., M.A (Columbia Univ.) Student at Columbia University, *21428-112 Road, Queens Village 9, New York.*

Bahadur, R. R., M.A. (Univ. of Delhi, India) Graduate Student at University of North Carolina, Chapel Hill, North Carolina.

Blom, Gunnar, Fil.kand. (Stockholm) *Olof Skotkonungs vag 8, Aspudden, Sweden*

Burrows, Glenn L., M.A. (Michigan State College) Research Associate, P.O. Box 168, Institute of Mathematical Statistics, Chapel Hill, North Carolina

Chapman, Carlos A., Jr., M S (Univ of Michigan) Sales Statistician, Argus, Inc., Ann Arbor, Michigan, *834 W. Huron St , Ann Arbor, Mich.*

Chiang, Chin Long, M.A (Univ of Calif.) Student at the University of California, *336-A Panoramic Way, Berkeley 4, California*

Coggins, Paul B., M S. (Univ of Wisconsin) Graduate Teaching Assistant, University of Michigan, *University Club, Madison 5, Wisconsin.*

Crapsey, Marcus T., A.B. (Univ. of Michigan) Graduate student at the University of Michigan, *615 Monroe, Ann Arbor, Michigan.*

Coy, John W., M A. (Univ. of New Mexico) Teaching Fellow, Department of Mathematics, University of Michigan, *2044 Whitewood, Ann Arbor, Michigan.*

Cutkosky, Richard E., Student at Carnegie Institute of Technology, *Box 401, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.*

DelPriore, Francis R., B.A. (New York Univ.) Associate Statistician, U S. Naval Engineering Experiment Station, *2609-22nd Street, N.E., Washington 18, D. C.*

Desind, Philip, M.S. (College of City of N. Y.) Statistician, Bureau of Ships, Navy Department, Washington, D. C , *7418 Georgia Ave., N.W., Washington, D. C.*

Dutka, Solomon, M A (Columbia Univ) Chief Statistician, % Elmo Roper, 30 Rockefeller Plaza, New York City, New York.

Dwass, Meyer, B.A. (George Washington Univ.) Graduate student at Columbia University, *Apt. 3A, 609 W. 115 St., New York, New York*

Eastman, Walter F., A.B. (Harvard) Central Technical Department, The American Brass Co , Waterbury, Connecticut.

Eisenpress, Harry, B.A (College of City of N. Y.) National Bureau of Economic Research, 1819 Broadway, New York 23, New York, *2935 Ocean Parkway, Brooklyn 24, New York.*

Fellows, Clifford Martin, B.S. (Boston Univ.) Assistant Instructor, Boston University, Bureau of Research and Statistics, *685 Commonwealth Avenue, Boston 15, Massachusetts.*

Gowen, John W., Ph.D. (Columbia Univ.) Professor of Genetics, Genetics Department, Iowa State College, *2014 Kildee, Ames, Iowa.*

Greenwood, Robert E., Ph.D. (Princeton Univ.) Assistant Professor of Applied Mathematics, University of Texas, *1704 Windsor Road, Austin, Texas*

Hald, Anders, Ph.D. (Univ. of Copenhagen) Professor of Statistics, University of Copenhagen, *Emdrupvenge 94, Copenhagen 0, Denmark*

Helms, William R., Student at Ohio State University, *Stadium Club, Ohio State University, Columbus 10, Ohio.*

Hemphill, F. M., M.S Ph. (Univ. of Michigan) Major, U. S. Public Health Service, School of Public Health, University of Michigan, Ann Arbor, Michigan.

Himes, Harold W., B.S. (George Pepperdine College, Los Angeles) Statistician, Test Design and Analysis Section, U.C D.W.R., U. S. Navy Electronics Laboratory, San Diego 52, California.

Hutchinson, L. Charles, Ph.D. (Mass. Institute of Tech.) Associate Professor of Mathematics, Polytechnic Institute of Brooklyn, Brooklyn, New York.

Klahr, Carl N., M S. (Carnegie Institute of Tech.) Student, Atomic Energy Commission Fellow, Carnegie Institute of Technology, *6337 Phillips Avenue, Pittsburgh 17, Pennsylvania*

Kraemer, Herbert F., B S (Univ of Delaware) Statistical Engineer, Technical Supervisor, Commercial Solvents Corporation, Terre Haute, Indiana, *1514 South 7th St., Terre Haute, Indiana.*

Kuebler, Roy R., Jr., A.M. (Univ of Pennsylvania) Associate Professor of Mathematics, Dickinson College, Carlisle, Pennsylvania.

Lafontant, Herne E., M S. (Atlantic Univ.) Student at the University of Michigan, *615 Monroe, Ann Arbor, Michigan.*

Lal, Dip Naravan, Ph D. (Edinburgh Univ.) Lecturer in Mathematics, Patna University, *New Dak Bungalow Road, Patna, Bihar, India.*

Liserre, Guido Orlando G., Profesor de Estadistica, *Mendoza 2540, Rosario, R., Argentina*

Matson, J. H., B A. (Univ. of Wisconsin) Statistician, Baker Manufacturing Company, Evansville, Wisconsin.

Monsch, Henry D., B.S. (Missouri School of Mines & Metallurgy, Rolla) Metallurgist, Aluminum Company of America, Fabricating Division, Alcoa, Tennessee, *5507 Lake Shore Drive, Knoxville, Tennessee*

Moore, Lucius T., Ph D. (John Hopkins Univ.) Associate Professor, Department of Mathematics, Brooklyn College, *205 Hicks Street, Brooklyn, New York.*

Noack, Albert, Ph.D. (Kiel, Germany) Privatdozent, Studienrat, (*24a*) *Hamburg-Lokstedt II, Tibarg 26, Germany.*

Patton, Robert E., A B. (N Y. State Teachers College, Albany) Graduate student at the University of Michigan, *522 Linden St., Ann Arbor, Michigan.*

Potter, Muriel, Ph.D. (Columbia Univ.) Instructor in Psychological Foundations, Educational Research and Reading Supervisor, Teachers College, Columbia University, *414 Riverside Drive, New York 25, New York.*

Putz, Robert R., B.A (Univ of Minnesota) Teaching Assistant, Department of Mathematics, University of California, *1631 Cornell Avenue, Berkeley 2, California.*

Ratoosh, Philburn, M.A. (Columbia Univ.) Assistant in Psychology, Department of Psychology, Columbia University, New York 27, New York

Richardson, Wyman, Jr., S.B. (Harvard) Graduate student at the University of North Carolina, *208-B, Chapel Hill, North Carolina.*

Rosenbaum, Sidney, M A. (Cambridge) Scientific Officer, Ministry of Works, 31, Multon House, Shore Place, London E.g., England

Savage, I. Richard, M.S. (Univ. of Michigan) Student at Columbia University, 1414 John Jay Hall, New York 27, New York.

Sheerin, Gail, A B (Univ. of Rochester) Statistical Technician, A.E.C. Project, University of Rochester, *1091 Highland Avenue, Rochester, New York.*

Siegert, Arnold J. F., Ph.D. (Leipzig, Germany) Professor of Physics, Department of Physics, Northwestern University, Evanston, Illinois

Simpson, Paul B., Ph.D. (Cornell Univ.) Assistant Professor of Economics, Department of Economics, Stanford University, California.

Solem, Anson D., M.S. (Harvard Univ.) Chief of Fragmentation Section, Naval Ordnance Laboratory, White Oak, Maryland, *121 Galveston St., S.W., Washington 20, D. C*

Sorensen, Frederick A., B S. (Carnegie Institute of Tech.) Teaching Assistant in Mathematics, Carnegie Institute of Technology, *1204 East End Avenue, Pittsburgh 18, Pennsylvania.*

Steel, Robert G. C., M.A. (Acadia Univ., Canada) Instructor and Research Associate, Statistical Laboratory, Iowa State College, Ames, Iowa

Taylor, Francis B., A.M. (Columbia Univ.) Instructor in Mathematics, Manhattan College, New York and Graduate student at Columbia University, *345 E. 193 St., Bronx 58, New York.*

Terrell, James R., A.B. (Univ. of Michigan) Statistical Clerk, Research Center for Group Dynamics, *P O Box 357, Ann Arbor, Michigan.*

Tick, Leo J., B.S. (Iowa State College) Research Graduate Assistant, Statistical Laboratory, Iowa State College, Ames, Iowa.

Tyler, Sylvanus A., S.M. (Univ. of Chicago) Associate Mathematician (Biometrics), Argonne National Laboratory, P.O. Box 5207, *9059 So. Stewart Avenue, Chicago 20, Illinois.*

Tysver, Joseph B., M.A. (Washington State College) Teaching Fellow, University of Michigan, *1404 Erving Court, Willow Run Village, Michigan.*

Umarji, Raghavendra R., A.M. (Columbia Univ.) Lecturer in Mathematics, Bombay Educational Service, 509 John Jay Hall, Columbia University, New York 27, New York.

Wilburn, A. J., A.B. (Howard Univ.) Statistician, Civil Aeronautics Board, Washington, D. C., *25-46th Place., N E., Washington, D. C.*

Correction

The information following Paul Koditschek's name which appeared in the March issue of the *Annals*, page 149, should have appeared as follows:

Koditschek, Paul, Ll. D. (Univ. of Vienna) Research Associate, Scientific Research Service, *319 W. 13th Street, New York 14, New York.*

(It was implied in the original notice that Scientific Research Service is connected with Columbia University.)

News Item from Cornell

With the continued support of a research contract with the Office of Naval Research, the Mathematics Department of Cornell University is further expanding research and instruction in the theory of probability and its applications. At present Professors Feller, Kac, Chung and Dr. Donsker are participating in the work. Professor G. Elfving of the University of Helsingfors has been appointed Visiting Professor of Mathematical Statistics for the academic years 1949-1951. Professor J. L. Doob, on sabbatical leave from the University of Illinois, will spend the year 1949-50 at Cornell. Dr. Gilbert Hunt has been appointed Assistant Professor of Mathematics.

REPORT ON THE NEW YORK MEETING OF THE INSTITUTE

The thirty-eighth meeting of the Institute of Mathematical Statistics was held at Columbia University, New York City on Friday afternoon and Saturday, April 8–9, 1949. The meeting was attended by 93 persons including the following 80 members of the Institute:

A. Abruzzi, T. W. Anderson, Leo A. Aroian, Robert Bechhofer, A. A. Bennett, Joseph Berkson, Allan Birnbaum, C. I. Bliss, Paul Boschan, P. G. Carlson, Uttam Chand, Yunien Chen, E. P. Coleman, T F Cope, Jerome Cornfield, L M Court, M. I Cropsen, J H. Curtiss, Cuthbert Daniel, F R Del Priore, W. E. Deming, J. A Dudman, David Durand, C. W. Dunnett, A Dvoretzky, P. S. Dwyer, Churchill Eisenhart, H. L Edgett, Harry Eisenpress, Lillian R. Elveback, D. A. S. Fraser, Murray Geisler, L. A. Goodman, J. I. Griffin, C C. Grove, E J. Gumbel, Miriam S. Harold, Mina Haskind, L H. Herbach, Harold Hotelling, Cuthbert Hurd, Arthur Kaufman, Roger D. Keeney, Paul Koditschek, Carl F Kossack, Howard Levene, Jack Laderman, I. D. Lorge, C. L Marks, Paul Meier, Frederick Mosteller, E B. Mundie, C. M Mottley, I. U Mulk, Paul Neurath, G. E. Noether, Doris Newman, M. L Norden, E. W. Pike, J. K Perrin, H. M. Rosenblatt, Frank Saidel, William Salkind, F. E Satterthwaite, Richard Savage, Henry Scheffé, H L. Seal, Jack Sherman, Rosedith Sitgreaves, J. H Smith, J J Sodano, Herbert Solomon, Mary N Torrey, J. W. Tukey, S. S Wilks, D. F. Votaw, Helen M Walker, Lionel Weiss, Jack Wolfowitz and W. W. Wryht.

The Friday afternoon session consisted of a Symposium on *Applications of Multivariate Analysis*, Professor S. S. Wilks of Princeton University presiding. The following two invited papers were given:

1. *Tests of Differences in Composite Growth Measurements in Pig Feeding Trials*, J. Wishart, Cambridge University and University of North Carolina.
2. *Fields of Application of Multivariate Analysis*, Harold Hotelling, University of North Carolina.

The prepared discussion was presented by Professor S N. Roy, Presidency College, Calcutta, and Columbia University, followed by discussion from the floor.

The Saturday morning sesssion was opened by a business meeting, Dr. Churchill Eisenhart, National Bureau of Standards, presiding. Among other items of business the Constitution of the Institute was amended to provide for Institutional Membership, and the by-laws amended to specify the status and privileges of Institutional Members. The revised Constitution and By-Laws appear elsewhere in this issue.

The second part of the session, Dr. W. Edwards Deming presiding, was devoted to an invited address: *Non-Linear Regression Laws and "Internal Least Squares,"* by Dr. H. O. Hartley, University College, London and Princeton University.

At the Saturday afternoon session, Professor Henry Scheffé, Columbia Uni-

versity, presiding, the following contributed papers were presented, ten in person, three by title:

1. *Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix.*
   Jack Sherman and Winifred J. Morrison, The Texas Company Research Laboratories, Beacon, N. Y.
2. *The Distribution of the Number of Exceedances.*
   E. J. Gumbel, New York, N. Y., and H. von Schelling, Naval Research Laboratory, New London, Connecticut.
3. *Note on the Power Function of a Quality Control Chart.*
   Leo A. Aroian, Hunter College
4. *Tests between Two Means or Regression Coefficients When Observations Are of Unequal Precision.*
   Uttam Chand, University of North Carolina.
5. *Functional Expansions.*
   Eugene W. Pike, Boston, Massachusetts.
6. *The Geometric Range for Distributions of Cauchy's Type.*
   E J Gumbel, New York, N Y., and R. D. Keeney, Metropolitan Life Insurance Company, New York, N. Y.
7. *On Sums of Random Integers Reduced Modulo m.*
   A. Dvoretzky, Hebrew University, Jerusalem, and Institute for Advanced Study, and J. Wolfowitz, Columbia University.
8. *The Corpuscle Problem: Estimating the Surface-Volume Ratio of a Corpuscle of Arbitrary Shape.*
   Jerome Cornfield, National Institute of Health, and Harold W. Chalkey, National Cancer Institute, Bethesda, Maryland.
9. *Generalized Hit Probabilities with a Gaussian Target.*
   D. A S. Fraser, Princeton University.
10. *A New Continuous Sampling Inspection Plan Based on an Analysis of Costs.*
    F E. Satterthwaite, General Electric Company, Bridgeport, Connecticut.
11. *On Levels of Significance of the F and Beta Distributions.* (By title)
    Leo A. Aroian, Hunter College.
12. *Certain Statistics for Samples of 3 from a Rectangular Distribution.* (By title)
    Julius Lieblein, Statistical Engineering Laboratory, National Bureau of Standards.
13. *The Choice of Lot Inspection Plans on the Basis of Cost.* (By title)
    F. E Satterthwaite and Burton Grad, General Electric Company, Bridgeport, Connecticut.

On Friday evening a dinner was held at the Men's Faculty Club.

S. B. LITTAUER
Assistant Secretary

CONSTITUTION OF THE INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE 1

Purpose

The Institute of Mathematical Statistics is a society for encouraging the development, dissemination, and application of mathematical statistics.

ARTICLE 2

Members

The Institute shall have Members and Institutional Members. Applications for membership must be approved by the Council. The Council may delegate this authority.

Except for nonpayment of dues, no Member or Institutional Member shall be expelled or suspended except by three-fourths vote of the Council.

ARTICLE 3

Officers

The Officers of the Institute shall be the President, the President-Elect, the Secretary, the Treasurer, and the Editor. The terms of office of the Secretary, the Treasurer and the Editor shall be three years. The terms of office of the President and the President-Elect shall be one year. The President-Elect shall succeed the President in that office. If the President is incapacitated, the President-Elect shall act as President, or, in case the President-Elect is also incapacitated the Secretary shall so act. Incapacity shall be determined by the Council.

The President shall act as chairman of the Council and of the Executive Committee, and shall appoint the Committees and representatives of the Institute, with the exception of the Committee on Fellows and the Executive Committee. Such Committee appointments shall be for terms of not more than three years, provided that committee appointments extending beyond the current year shall be either to standing committees with regularly rotating membership, or to temporary committees assigned specific tasks.

The Treasurer shall present financial statements to the Council and shall bring condensed statements to the attention of the Members.

The Secretary shall record the actions of the Council and of the Executive Committee and of Institute meetings, arrange for and inform the Members of meetings and conduct the correspondence of the Institute except as otherwise assigned by the Executive Committee. The Secretary may appoint Assistant Secretaries to assist him in connection with specified meetings or for other occasions. The offices of Secretary and Treasurer may be combined.

ARTICLE 4

Council

The Council shall consist of not less than twelve elected members in addition to the Officers of the Institute except that vacancies in the Council occurring subsequent to an election shall not be filled until the next annual election.

Elected members shall be elected for terms of three years, the terms of approximately one-third of them terminating each year.

The Council, representing the Members, shall determine the policies and supervise the affairs of the Institute in accordance with any Bylaws the Institute may adopt. It shall determine the standing committees of the Institute and the number of elected members of the Council.

The Council shall elect the Secretary, the Treasurer, and the Editor, by majority vote The Council shall determine the number, if any, of Associate Secretaries, Associate Treasurers and Associate Editors. The Secretary shall nominate Associate Secretaries, the Treasurer shall nominate Associate Treasurers, and the Editor shall nominate Associate Editors which the Council may elect by majority vote. Such Associate Secretaries, Treasurers, and Editors shall be non-voting members of the Council.

The Council shall meet at least twice a year, usually at times of meetings of the Institute, and otherwise at the call of the President or the call of any five members of the Council. Any voting member unable to be present may appoint, in writing, a representative to speak for him, and such representative shall be entitled to vote. A quorum shall be seven persons entitled to vote. Majorities and other fractions of the Council are to be based on the number of persons present and entitled to vote

ARTICLE 5

Executive Committee

The Officers shall constitute the Executive Committee of the Council, and shall conduct the affairs of the Institute.

The Executive Committee may create temporary committees with assigned tasks coming within the scope of the Institute.

ARTICLE 6

Nominations

The President shall appoint a Nominating Committee and shall announce their names at the annual meeting when he retires from office. This Committee shall submit to the Members, through the Secretary and at least sixty days before the closing of polls at the next succeeding annual meeting, one nomination for President-Elect and a slate containing at least twice as many names as there are vacancies on the Council.

Additional nominations may be made for President-Elect or for the Council by a petition signed by twenty Members. Such nominations shall appear on

the ballot if they are in the hands of the Secretary at least 30 days before the closing of polls at the next succeeding annual meeting. In any event, Members may vote for names in addition to those nominated.

ARTICLE 7

Fellows

The Council, may, by majority vote, elect to fellowship any Member nominated by the Committee on Fellows. Such nomination and election shall be on the basis of the nominee's contributions to the development, dissemination, and application of mathematical statistics.

ARTICLE 8

Committee on Fellows

The Council shall elect two Fellows annually to serve for three years on the Committee on Fellows. One of the Members whose term is next to expire shall be designated by the President as chairman.

ARTICLE 9

Publications

The Annals of Mathematical Statistics shall be the official journal of the Institute. Other publications may be authorized by the Council.

The publications of the Institute shall be supervised by the Editor, with the assistance of the Associate Editors and such committees as the Council may approve.

ARTICLE 10

Communications

Public announcements concerning the Institute, including statements of policy, recommendations, reports of committees and accounts of Council meetings shall be issued by the Secretary or the President with the prior approval of the Council or its Executive Committee. Advance publicity concerning meetings may be released by authorized Program Committees or Publicity Committees

ARTICLE 11

Affiliation

By a three-fourths vote, the Council may authorize the affiliation of the Institute with any organization whose aims are consistent with those of the Institute.

ARTICLE 12

Amendments

This constitution may be amended by an affirmative two-thirds vote of those Members voting at any regularly convened meeting of the Institute provided

notice of such proposed amendment shall have been sent to each Member by the Secretary at least thirty days before the date of the meeting at which the proposal is to be acted upon. Members may vote in person or by mail. The Secretary shall send to the Members any amendments recommended by the Executive Committee or proposed through a petition of 25 members of the Institute.

ARTICLE 13

Emergencies

In an emergency, as determined by the President or the Executive Committee, or by a majority of the Council, a meeting of the Council to transact business or a meeting of the Institute to amend the constitution may be conducted by mail.

BY-LAWS OF THE INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE 1

Duties of Officers

The President, or in his absence the President-Elect, or in his absence a Member appointed by the Executive Committee, shall preside at business meetings of the Institute.

The Treasurer shall send out calls for annual dues, pay all bills for expenditures authorized by the Institute, Council, or Executive Committee; keep a detailed account of all receipts and expenditures; prepare a financial statement at the end of each fiscal year and present an abstract of same at a business meeting of the Institute after it has been audited by a Member or Members appointed by the President, to whom such Member or Members shall report.

The Secretary shall, subject to the direction of the Council, have charge of the archives and other tangible and intangible property of the Institute and shall, upon the direction of the Council, publish a classified list of all Members of the Institute, and of Institutional Members at their request.

The Editor, subject to the direction of the Council, shall have charge of all editorial matters, whether relating to the official Journal or to other publications. He shall, with the advice and consent of the Council, appoint an Editorial Committee of not less than twelve Members to cooperate with him for definite terms. All appointments to the Editorial Committee shall terminate with the appointment of a new Editor.

ARTICLE 2

Dues

Members shall pay seven dollars at the time of admission to membership and shall receive the full current volume of the official Journal. Thereafter Members shall pay seven dollars annual dues, of which five dollars shall be for a subscription to the Official Journal There shall be the following exceptions:

A. Two Members of the Institute who are husband and wife may elect to receive one copy of the Official Journal between them, when their dues shall each be reduced by twenty-five percent.
B. Any Member may make a payment in place of all succeeding annual dues based on a suitable table and rate of interest specified by the Council.
C. Any Member on active military duty may notify the Treasurer that he wishes neither to pay dues nor to receive the Official Journal during the current year. He may receive the official Journal for the suspended years on payment of one-half of the suspended dues within one year after resuming payment of annual dues.
D. Any Member who resides outside the Western Hemisphere shall pay five dollars annual dues.

Institutional Members shall pay annual dues of at least $100. For each $100 of annual dues, an Institutional Member shall receive two copies of the Official Journal, one bound, and shall be entitled to designate one person to have the full prerogatives of a member without further payment of dues (including the receipt of a personal copy of the Official Journal). Twenty-five dollars of each $100 shall be allocated to the three subscriptions to the Official Journal and the binding of one copy.

Annual dues shall be payable on the first day of January of each year.

It shall be the duty of the Treasurer to notify by mail anyone whose dues are six months in arrears, enclosing a copy of this article. If such person fails to pay such dues within three months from the date of mailing such notice, the Treasurer shall report the delinquent to the Council, who may suspend the delinquent from membership and who may reinstate the delinquent upon payment of arrears.

ARTICLE 3

Salaries

The Institute shall not pay a salary to any Officer, Councilor, or member of any committee.

ARTICLE 4

Amendments

These Bylaws may be amended in the same manner as the Constitution or, if the proposed amendment has been previously approved by the Council, by a majority vote at any regularly convened meeting.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

DECEMBER, 1948

Articles

Commercial Uses of Sampling — J. Stevens Stock and Joseph R. Hochstim
Variation of the Frequency of Fatal Quarrels with Magnitude — Lewis F. Richardson
Bank Reserves and Business Fluctuations — Clark Warburton
The Ordering of n Items Assigned to k Rank Categories by Votes of m Individuals — Garret L. Schuyler
Levels of Significance for Variance Ratio of Two Samples of Equal Size — C. J. Kirchen
Main Effects and Interactions — D. J. Finney
A Test for Symmetry in Contingency Tables — Albert H. Bowker
The War Production Board's Statistical Reporting Experience, Part IV — David Novick and George A. Steiner
Correction to "On Estimating Precision of Measuring Instruments and Product Variability" — Frank E. Grubbs
Statistical Methodology Index — Oscar Krisen Buros

AMERICAN STATISTICAL ASSOCIATION
1603 K Street, N. W., Washington 6, D. C.

MATHEMATICAL REVIEWS

A journal containing reviews of the mathematical literature of the world, with full subject and author indices

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, Institute of Mathematical Statistics, London Mathematical Society, Edinburgh Mathematical Society, Union Matematica Argentina, and others

Subscriptions accepted to cover the calendar year only.
Issues appear monthly except July. $20.00 per year.

Send subscription order or request for sample copy to

AMERICAN MATHEMATICAL SOCIETY
531 West 116th Street, New York City 27

ON THE THEORY OF SYSTEMATIC SAMPLING, II

By William G. Madow[1]

Institute of Statistics, University of North Carolina

1. Summary and introduction. In an earlier paper,[2] [1] an approach to the problem of systematic sampling was formulated, and the associated variance obtained. Several forms of the population were assumed. The efficiency of the systematic design as compared with the random and stratified random design was evaluated for these forms. It was remarked that as the size of sample increased the variance of a systematic design might also increase, contrary to the behavior of variances in the random sampling design This possibility was verified in [2].

One approach to the study of systematic designs, given by Cochran [3] removed this difficulty to some extent by changing the problem to one of the expected variance, and supposing the elements of the population to be random variables. He showed that if the correlogram of these random variables is concave upwards, then the expected variance of the systematic design would be less, and often considerably less, than the variance of a stratified design.

In the present paper the results of the earlier papers are extended to the systematic sampling of clusters of equal and unequal sizes. Some comments on systematic sampling in two dimensions are included.

In section 2 we derive two theorems that have considerable applications in many parts of sampling. Although it has been common for people working in sampling theory to tell each other that these theorems ought to be true, yet no reference seems to exist.

In section 3 we develop the implications of a remark [1, p 13] that in designing sample surveys we should try to induce negative correlation between strata. In Theorem 3 we obtain sufficient conditions for the correlation to be negative. The lemma and Theorem 4 given in Section 4 enable us to extend the uses of Theorem 3 in practice As an application of these results, we show that if a population has a concave upwards correlogram, and if strata are defined in an optimum fashion for the selection of one element at random from each stratum, then we can define a systematic type design that will be more efficient than independent random selection from each stratum.

In sections 5 and 6 we obtain various results in the systematic sampling of clusters largely as applications of the more general theorems of the earlier sections. In general the results are of a nature similar to those of [1] and [3] in that the formulae show the conditions under which systematic sampling may be expected to be more efficient than random or stratified random sampling We have not, however, applied these formulae to specified types of populations

[1] Submitted for publication, November, 1948. Parts of this paper were prepared while the author was Visiting Professor of Statistics at the University of São Paulo, Brazil

[2] References to the articles and book cited are given by Roman numerals

From [1, 2 and 3] it is already apparent that this work will be useful and such studies should be more valuable when made in connection with important types of surveys or data than when made as illustrations in a general paper.

2. Random events and conditional expectations. Almost invariably, samples are selected in several stages. For example, to select a sample of households from a city one frequently used method is the following two stage sampling plan:

a. A map of the city showing the location of each block is obtained and brought up-to-date.
b. Using this map, a sample of the blocks of the city is selected (this is stage1).
c. From the households on the blocks selected in stage 1, a subsample of households is selected (this is stage 2.).

In this section, we give a general approach for evaluating the means and variances associated with multi-stage sampling. This approach has the advantage of at once yielding the contributions to the variance arising from each stage. Furthermore, the theorems presented are useful in calculating variances even when our interest is not in multi-stage sampling. The theorems are presented in general terms because of their wide application in sampling.

We shall say that the result of performing an operation is a random event A^* if the result can assume m possible states $A_1, \cdots, A_m$ with probabilities $p_1, \cdots, p_m$, where

$$P\{A^* = A_i\} = p_i, \qquad \sum_{i=1}^{m} p_i = 1,$$

and $P\{A^* = A_i\}$ is read "the probability that the random event A^* assumes the state A_i."

One illustration of an operation is the operation of selecting a sample of blocks. If there are N blocks in the city of which we select n in such a way that each set of n of the N blocks is a possible sample, then there are C_n^N possible samples. In this case $m = C_n^N$ and the C_n^N possible samples are the m states of A^* "the result of selecting the sample of blocks." Furthermore, if each of the possible samples of blocks is equally likely to be selected, then

$$P\{A^* = A_i\} = \frac{1}{C_n^N} = \frac{1}{m}.$$

The random event A^* may also be the taking on by a random variable of one of its possible values. If z^* is a random variable having possible values $z_1, \cdots, z_m$ with probabilities $p_1, \cdots, p_m$ then we can define the states of A^* to be A_i where A_i is "$z^* = z_i$."

Thus the notion of a random event includes the two types of randomness that are met in selecting samples.

Let x' be a random variable. Then, by the conditional expectation of x' subject to the random event A^* is meant the random variable $E^*(x' \mid A)$ whose possible values are $E(x' \mid A_i)$, $i = 1, \cdots, m$ and whose probabilities are p_i, that is

$$P\{E^*(x' \mid A) = E(x' \mid A_i)\} = p_i = P\{A^* = A_i\},$$

where

$$E(x' \mid A_i) = \sum_{j=1}^{N_i} x_{ij}\, p_j(A_i), \tag{2.1}$$

x_{ij} is the jth of the N_i possible values of x' when A_i occurs, and

$$p_j(A_i) = P\{x' = x_{ij} \mid A_i\}$$

is "the probability that $x' = x_{ij}$ given that A_i occurs." It should be noted that if

$$p_{ij} = P\{x' = x_{ij}\},$$

then

$$p_{ij} = P\{x' = x_{ij}, A^* = A_i\}$$

since the fact that $x' = x_{ij}$ implies the occurrence of A_i. Then

$$p_i \cdot p_j(A_i) = p_{ij}. \tag{2.2}$$

We state Theorems 1 and 2 without proof since their proofs are immediate.

THEOREM 1. *The expected value of the random variable $E^*(x' \mid A)$ is $E\,x'$*, i.e.

$$E\{E^*(x' \mid A)\} = Ex'.$$

By $\sigma^*_{x'y'|A}$ we shall mean the random variable whose possible values are $\sigma_{x'y'|A_i}$, $i = 1, \cdots, m$ where

$$\sigma_{x'y'|A_i} = E\{[x' - E(x' \mid A_i)]\,[y' - E(y' \mid A_i)] \mid A_i\}$$

and

$$P\{\sigma_{x'y'|A} = \sigma_{x'y'|A'}\} = p_i = P\{A^* = A_i\},$$

i.e.

$$\sigma^*_{x'y'|A} = E^*\{[x' - E^*(x' \mid A)]\,[y' - E^*(y' \mid A)] \mid A\}.$$

Furthermore, the symbol $\sigma_{E^*(x'|A)\,E^*(y'|A)}$ will stand for "the covariance of the two random variables $E^*(x' \mid A)$ and $E^*(y' \mid A)$" The corresponding definitions of variance are obtained by replacing y' by x' above.

THEOREM 2. *If x' and y' are random variables, then*

$$\sigma_{x'y'} = E\sigma^*_{x'y'|A} + \sigma_{E^*(x'|A)\,E^*(y'|A)}$$

and

$$\sigma^2_{x'} = E\sigma^2_{x'|A} + \sigma^2_{E^*(x'|A)}.$$

We note that, since the p_{ij}, p_i and $p_j(A_i)$ are not specified, Theorems 1 and 2 are valid for any two-stage plan. The generalizations of Theorems 1 and 2 to multi-stage plans are obvious, but in practice it often turns out to be simpler to apply the theorems several times.

It would be easy to give applications of Theorems 1 and 2 but these are not essential for our purposes in this paper. As remarked in the introduction, these two theorems have long been part of what we may call the folklore of sampling

3. Stratified sampling and negative correlation, with an application to systematic sampling. In discussing plans for sampling from a stratified population it is customary to suppose that if x' is an estimate and $x' = x_1' + \cdots + x_L'$ where x_j' is the contribution to x' arising from the jth of the L strata, then the sampling is to be so done that the random variables x_i' and x_j', $j \neq i$, are independent

In [1, p. 13] it was noted that if a population were stratified, and if the elements were so selected that the contributions from different strata were negatively correlated, it would follow that the variance of the estimate would be less than if the contributions were independent but had the same covariances within strata. This was, of course, an immediate conclusion from the fact that

$$\sigma_{x'}^2 = \sum_{i,j=1}^{L} \sigma_{x_i' x_j'}$$

and, hence, if

$$C = \sum_{i \neq j} \sigma_{x_i' x_j'} < 0 \tag{3.1}$$

then $\sigma_{x'}^2$ is less than it would be if $C = 0$. If $C < 0$ we shall say that the sample design has "negative correlation."

It is obvious that any population may be taken to be itself a sample, a sample from the possible populations that might have been produced by the forces that determined the existing population. Inasmuch as sampling designs are often chosen on the basis of a knowledge of the dominating forces and some past experience, it is realistic to consider not only the expected values and variances for a specific population but also their expected values over all possible populations determined by the same forces. Cochran [3] has given one illustration of the usefulness of considering the expected variance of a sample design. He considered the elements $x_1, \cdots, x_n$ of the population themselves to be random variables and supposed that $E x_i = \mu$ and $E(x_i - \mu)^2 = \sigma^2$. For his purposes it was also convenient to suppose that if $u > 0$ then $E(x_i - \mu)(x_{i+u} - \mu) = \rho_u \sigma^2$. It was then possible for him to make realistic hypotheses concerning the correlogram, i.e. the ρ_u considered as a function of u, that would not have been reasonable in dealing with a specific population. He thus obtained general conclusions concerning the expected efficiency of systematic sampling designs as compared with random and stratified random designs

In this paper we shall consider not only the expected values and variances for the given finite population but also the expected values of these expected values and variances under the assumption that the elements of the population are themselves random variables. We shall use $\mathfrak{E}$ to denote the expected value

considering the elements of the population to be random variables and as before use E for expected values based on the specified finite population.

Then

$$\mathfrak{E}\sigma^2_{x'} = \sum_{i,j=1}^{L} \mathfrak{E}\sigma_{x_i' x_j'},$$

and if $\mathfrak{E}C < 0$ we shall say that the design has 'expected negative correlation'

We now propose to obtain the beginnings of an approach to sample design when it is possible to introduce or take advantage of negative correlation or expected negative correlation through the sample design.

To simplify, we shall begin by considering two strata and shall suppose that the possible values of x' are $x_1, \cdots, x_n$ while the possible values of y' are $y_1, \cdots, y_n$. Furthermore, we shall suppose the sampling to be so done that

$$P\{x' = x_i\} = P\{y' = y_i\} = P\{x' = x_i, y' = y_i\} = p_i > 0,$$

so that $\sum_{i=1}^{n} p_i = 1$ and $P\{x' = x_i, y' = y_j\} = 0$ if $i \neq j$.

Under the above assumptions, it follows that

$$\sigma_{x'y'} = \sum_{i=1}^{n} p_i x_i y_i - \sum_{i,j=1}^{n} p_i p_j x_i y_j . \tag{3.2}$$

The symbol $\varphi_{ij} \geq 0$ means that $\varphi_{ij} \geq 0$ for all i and j and $\varphi_{ij} > 0$ for at least one pair i, j. We shall say that if $(x_i - x_j)(y_i - y_j) \geq 0$ then the sets (x) and (y), where (x) stands for $x_1, \cdots, x_n$ and (y) for $y_1, \cdots, y_n$ are similarly ordered and if $(x_i - x_j)(y_i - y_j) \leq 0$ then these sets are oppositely ordered. Then it is easy to prove, [4, p. 43] directly that if the values are oppositely ordered, then $\sigma_{x'y'} < 0$ and if they are similarly ordered then $\sigma_{x'y'} > 0$.

A somewhat more general result is the following:

THEOREM 3. *Let $n \leq k$, let*

$$b = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij} w_i z_j$$

be a real bilinear form, and let

$$t = \sum_{i=j}^{n} a_{ii} w_i$$

be a real linear form, where $w_i > 0$, $z_i > 0$ and $\sum_{i=1}^{n} w_i = \sum_{i=1}^{k} z_i = 1$.

Then a sufficient condition that $b > t$ is

$$a_{ij} \geq a_{ii} . \tag{3.3}$$

If $k = n$ and $w_i = z_i$ then $b > t$ if

$$a_{ij} + a_{ji} \geq a_{ii} + a_{jj} . \tag{3.4}$$

PROOF. Since

$$b - t = \sum_{i=1}^{n} a_{ii}(w_i z_i - w_i) + \sum_{i \neq j} a_{ij} w_i z_j ,$$

and since

$$1 - z_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} z_j ,$$

it follows that

$$b - t \sum_{j \neq i} (a_{ij} - a_{ii}) w_i z_j .$$

Hence, $b > t$ if (3.3) holds. Also, if $k = n$ and $w_i = z_i$ then $b > t$ if (3.4) holds. Some obvious generalizations of Theorem 3 have been omitted since we do not need them.

To obtain the result that $\sigma_{x'y'} < 0$ if the sets (x) and (y) are oppositely ordered, we make the identifications $a_{ij} = x_i y_j$ and $z_i = w_i = p_i$. Then (3.4) holds and substituting we have

$$a_{ii} + a_{jj} - a_{ij} - a_{ji} = (x_i - x_j)(y_i - y_j) \tag{3.5}$$

so that if the values are oppositely ordered, $\sigma_{x'y'} < 0$, and hence the two strata have negative correlation.

To consider expected negative correlation we note that

$$\mathcal{E}\sigma_{x'y'} = \sum_{i=1}^{n} p_i \sigma_{ii} + \sum_{i,j=1}^{n} p_i p_j \sigma_{ij} \tag{3.6}$$

where we suppose that $\mathcal{E}x_i = \mu$, $\mathcal{E}y_i = \nu$ and

$$\mathcal{E}(x_i - \mu)(y_j - \nu) = \sigma_{ij}$$

so that in this case σ_{ii} is a covariance, not a variance.

If we put $a_{ij} = \sigma_{ij}$ and $z_i = w_i = p_i$, then (3.4) holds and we obtain, as sufficient for $\mathcal{E}\sigma_{x'y'}$ to be negative, that

$$\sigma_{ij} + \sigma_{ji} \geq \sigma_{ii} + \sigma_{jj} \tag{3.7}$$

or, if we define ρ_{ij} by the equation,

$$\sigma_x \sigma_y \rho_{ij} = \sigma_{ij} ,$$

where $\sigma_x^2 = \mathcal{E}(x_i - \mu)^2$ and $\sigma_y^2 = \mathcal{E}(y_i - \nu)^2$, we have

$$\rho_{ij} + \rho_{ji} \geq \rho_{ii} + \rho_{jj} \tag{3.8}$$

as a sufficient condition for $\mathcal{E}\sigma_{x'y'} < 0$.

Let us consider the systematic sampling of single elements. In systematic sampling, we assume a population of kn ordered elements $x_1, x_2, \cdots, x_k, x_{1+k}, \cdots, x_{2k}, \cdots, x_{1+(n-1)k}, \cdots, x_{nk}$ of which we wish to estimate the arith-

metic mean $\bar{x}$. As our estimate we use

$$\bar{x}' = (x_1' + \cdots + x_m')/n$$

where x_1' is selected at random from $x_1, \cdots, x_k$ and if $x_1' = x_j$ then $x_i' = x_{j+(i-1)k}$, $i = 2, \cdots, m$. Thus, $\bar{x}'$ may be interpreted as an estimate based on a stratified population, the ith stratum consisting of

$$x_{1+(i-1)k}, \cdots, x_{k+(i-1)k}$$

and

$$P\{x_i' = x_{\alpha+(i-1)k}\} = P\{x_i' = x_{\alpha+(i-1)k}, x_j' = x_{\alpha+(j-1)k}\} = 1/k$$

while

$$P\{x_i' = x_{\alpha+(i-1)k}, x_j' = x_{\beta+(j-1)k}\} = 0, \text{ if } \alpha \neq \beta.$$

Then

$$\sigma_{x_i'x_j'} = \left(\frac{1}{k}\right) \sum_{\alpha=1}^{k} x_{\alpha+(i-1)k} \cdot x_{\alpha+(j-1)k} - \bar{x}_i \bar{x}_j$$

where

$$\bar{x}_i = \left(\frac{1}{k}\right) \sum_{\alpha=1}^{k} x_{\alpha+(i-1)k}.$$

Hence, any two strata that are oppositely ordered will yield a negative contribution to the variance. However, since it is not possible for all strata to be negatively ordered, we do not thus obtain a useful result and must return to the consideration of C or $\sigma_{\bar{x}}^2$ itself as was done in [1]. If, however, we make Cochran's assumptions, and consider $\mathfrak{E}\sigma_{x'y'}$, it follows that for the ith and jth strata

$$\rho_{\alpha\beta} = \rho_{(j-i)k+\beta-\alpha},$$

and (3.8) becomes

$$\rho_{(j-i)k+(\beta-\alpha)} + \rho_{(j-i)k+(\alpha-\beta)} \geq 2\rho_{(j-i)k}, \tag{3.9}$$

i.e. the correlation function ρ_u must be concave upwards, which Cochran showed by other means. By considering $\mathfrak{E}C$ it is possible to show that a sort of average concavity is all that is required of the correlogram for systematic sampling to have a smaller variance than stratified random sampling.

4. Conditions for negative correlation when the strata are of unequal sizes with an application to systematic sampling. Often, as in the systematic selection of clusters with probability proportionate to size (discussed in Section 5) the simplified situation dealt with in Theorem 3 does not directly apply. However, Theorem 3 may be used to advantage by the following device.

Let us suppose the possible values of x' to be $x_1, \cdots, x_n$ and those of y_0' to be $y_1^0, \cdots, y_k^0$, $k > n$ and let

$$P\{y' = y_\beta^0 \mid x' = x_\alpha\} = p_{\beta|\alpha}$$

so that if we define

$$y_\alpha = \sum_{\beta=1}^{k} y_\beta^0 p_{\beta|\alpha}, \tag{4.1}$$

then

$$y_\alpha = E(y_0' \mid x' = x_\alpha).$$

If we define y' to be a random variable having possible values $y_1, \cdots, y_n$ with probabilities $p_1, \cdots, p_n$ where

$$p_\alpha = P\{x' = x_\alpha\}$$

it follows that

$$y' = E^*(y_0' \mid x')$$

and

$$\sigma_{x'y_0'} = \sigma_{x'y'}.$$

Clearly, Theorem 3 is valid for the random variables x' and y'.

Consequently, we need only determine what restrictions the conditional probabilities, $p_{\beta|\alpha}$, and the values, y_α^0, need satisfy for the sets $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ to be oppositely ordered or for (3.7) to hold.

Substituting for y_i and y_j in (3.5) we see that if

$$(x_\alpha - x_\gamma) \sum_{\beta=1}^{k} y_\beta^0 (p_{\beta|\alpha} - p_{\beta|\gamma}) \leqq 0 \tag{4.2}$$

then $\sigma_{x'y'} = \sigma_{x'y_0'} < 0$.

Let

$$\sigma_{\alpha\gamma}^0 = \mathfrak{E}(x_\alpha - \mu)\,(y_\gamma^0 - \nu).$$

Then substituting in (3.7) we see that if

$$\sum_{\beta=1}^{k} (p_{\beta|\alpha} - p_{\beta|\gamma})(\sigma_{\alpha\beta}^0 - \sigma_{\gamma\beta}^0) \leqq 0 \tag{4.3}$$

or if

$$\sum_{\beta=1}^{k} (p_{\beta|\alpha} - p_{\beta|\gamma})(\rho_{\alpha\beta}^0 - \rho_{\gamma\beta}^0) \leqq 0 \tag{4.4}$$

then

$$\mathfrak{E}\sigma_{x'y_0'} < 0.$$

In order to use (4.2) and (4.3) the following well-known lemma is often useful.

LEMMA. *If* $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_k \leq 0$ *and the quantities* $\epsilon_1, \cdots, \epsilon_k$ *are such that*

$$\sum_{\beta=1}^{s} \epsilon_\beta \geq 0$$

then

$$\sum_{\beta=1}^{s} \epsilon_\beta \xi_\beta \leq 0, \qquad s = 1, \cdots, k.$$

Let us use this lemma to obtain another theorem that will be helpful in showing negative or expected negative correlation between strata.

THEOREM 4. *Let b be a bilinear form*

$$b = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} w_i z_j$$

such that $\sum_{i=1}^{s} w_i \geq 0$, $\sum_{j=1}^{s'} z_j \geq 0$, $s = 1, \cdots, n-1$, $s' = 1, \cdots, m-1$, *and*

$$\sum_{i=1}^{n} w_i = \sum_{j=1}^{m} z_j = 0 \tag{4.5}$$

Let

$$\delta_{ij} = a_{ij} - a_{i+1,j} - a_{i,j+1} + a_{i+1,j+1}.$$

Then a sufficient condition that $b \leq 0$ is $\delta_{ij} \leq 0$.

PROOF. Upon substituting for w_n and z_m in b from (4.5) we see that

$$b = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} \delta'_{ij} w_i z_j$$

where

$$\delta'_{ij} = a_{ij} - a_{im} - a_{nj} + a_{nm}$$

or, if we define,

$$\xi_j = \sum_{i=1}^{n-1} \delta'_{ij} w_i$$

then

$$b = \sum_{j=1}^{m-1} \xi_j z_j.$$

According to the lemma, it then follows that a sufficient condition that $b \leq 0$ is that

$$\xi_1 \leq \xi_2 \leq \cdots \leq \xi_{m-1} \leq 0.$$

Also, a sufficient condition that

$$\xi_j - \xi_{j+1} \leq 0,$$

is

$$\delta'_{ij} - \delta'_{i,j+1} \leq \delta'_{i+1,j} - \delta'_{i+1,j+1}$$

Then to complete the proof it is only necessary to verify that

$$\delta_{ij} = \delta'_{ij} - \delta'_{i,j+1} - \delta'_{i+1,j} + \delta'_{i+1,j+1} .$$

In the preceding pages we have given an identification of systematic with stratified sampling where, instead of the selection being made independently within strata, the choice of an element from one stratum determines the choice from the other strata. In this identification, however, it was assumed that the strata contained the same number of elements. Let us now extend this method of selecting samples to the case where the strata have different numbers of elements. In so doing we shall illustrate the use of the above lemma and theorem 4.

Suppose now that the population consists of N elements $x_1 , \cdots , x_N$ classified into n strata, the ith of which contains the N_i elements

$$x_{N_1+\cdots+N_{i-1}+1} , \cdots , x_{N_1+\cdots+N_i} .$$

We shall denote these elements by $x_{i1} , \cdots , x_{iN_i}$.

We shall select one element from each of these n strata. The element selected from the ith stratum is written x'_i . As the estimate of $\bar{x}$, the arithmetic mean of the population, we use

$$\bar{x}' = \sum_{i=1}^{n} \frac{N_i}{N} x'_i$$

and it is well known that if the selection is made independently at random from each stratum, then

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma_i^2$$

where σ_i^2 is the variance of x'_1 , i.e the variance of the ith stratum.

Let us now consider an alternative to the usual method. We can suppose that $N_1 > 1$ without any loss of generality. (The methods are the same for any stratum having $N_i = 1$ and will also yield the same result for any population such that either all the $N_i = 1$ or all but one of the $N_i = 1$. Differences occur if at least two of the N_i differ from 1.)

We first choose an element at random from the first stratum. Suppose that $x'_1 = x_\alpha$. Then to choose an element from the second stratum, assuming that $N_2 > 1$, we proceed as follows: Multiply N_2 by any positive integer t_2 such that $N_2 t_2 / N_1$ is an integer, say, k_2 . Assign to each element of the second stratum the measure of size t_2 , and form the two sets of cumulative totals $t_2 , 2t_2 , \cdots , N_2 t_2$ and $k_2 , 2k_2 , \cdots , N_1 k_2$. Then with the measures of size t_2 assigned to each element of stratum 2, and the measure of size k_2 assigned to each element of stratum 1, it follows that strata 1 and 2 have the same total size.

As an example of the arithmetic given below consider the following simple case. Suppose that $N_1 = 3$ and $N_2 = 4$. Then if we take for t_2 the value 6, it follows that $k_2 = 8$. We choose one of the integers 1, 2, 3 with equal probability. If the

integer 1 is obtained, we have selected the first element of the first stratum and choose an integer between 1 and 8 with equal probability. If the selected integer is between 1 and 6, the first element of the second stratum is selected. If it is 7 or 8 the second element of the second stratum is selected. Similarly if the second element of the first stratum is selected, then we select an integer between 9 and 16 with equal probability. If that integer has value 9, $\cdots$, 12 the second element of the second stratum is selected; if it has value 13, $\cdots$, 16 the third element is selected.

The general formulation of the selection procedure for the second stratum is:

Suppose that β_0 is the smallest integer such that $(\alpha - 1)k_2 + 1 \leq \beta_0 t_2$ and that β_1 is such that $(\beta_1 - 1)t_2 < \alpha k_2 \leq \beta_1 t_2$. Choosen an integer at random from $1, \cdots, k_2$ and call that integer β. Then, if

$$(\alpha - 1)k_2 < (\alpha - 1)k_2 + \beta \leq \beta_0 t_2$$

the β_0th element is selected from stratum 2; if

$$\beta_0 t_2 < (\alpha - 1)k_2 + \beta \leq (\beta_0 + 1)t_2$$

the $(\beta_0 + 1)$th element is selected; $\cdots$; and if

$$(\beta_1 - 1)t_2 < (\alpha - 1)k_2 + \beta \leq \alpha k_2$$

the β_1th element is selected from stratum 2.

It is easy to verify that when the sample is so selected, each element of stratum 2 has equal probability of being selected. Hence, if we apply this procedure to each stratum we have

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma_i^2 + \sum_{i \neq j} \frac{N_i N_j}{N^2} \sigma_{x_i' x_j'}.$$

Let us evaluate $\sigma_{x_i' x_j'}$ for this type of selection. Now

$$\sigma_{x_i' x_j'} = E\,(x_i' - \bar{x}_i)(x_j' - \bar{x}_j)$$

where $\bar{x}_i$ is the arithmetic mean of the elements of the ith stratum. From Theorem 2, we then have

$$\sigma_{x_i' x_j'} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} E[x_i' - E(x_i' \mid x_{1\alpha})][x_j' - E(x_j' \mid x_{1\alpha}]$$

$$+ \frac{1}{N_1} \sum_{\alpha=1}^{N_1} [E(x_i' \mid x_{1\alpha}) - \bar{x}_i][E(x_j' \mid x_{1\alpha}) - \bar{x}_j].$$

It is easy to see that the method of selection used above implies that the first term of $\sigma_{x_i' x_j'}$ vanishes. Furthermore, $\bar{x}_i$ is the arithmetic mean of the conditional expectations so that we have reduced the problem to one of determining whether the conditional expectations satisfy the conditions for negative correlation or expected negative correlation.

If we denote $E\,(x_i' \mid x_{1\alpha})$ by $y_{i\alpha}$, then we need to see whether the sets $y_{i1} \cdots$,

$y_{i N_1}$ and $y_{j_1}, \cdots, y_{j N_1}$ are oppositely ordered. Now

$$(y_{i\alpha} - y_{i\beta})(y_{j\alpha} - y_{j\beta} = \sum_{g=1}^{N_i} \sum_{h=1}^{N_j} x_{ig} x_{jh} \epsilon_{igh\alpha\beta} \epsilon_{jh\alpha\beta}$$

where

$$\epsilon_{igh\alpha\beta} = P\{x_i' = x_{ig} \mid x_{i\alpha}\} - P\{x_i' = x_{ig} \mid x_{i\beta}\}.$$

If $\alpha < \beta$ then, according to the method of selection,

$$\sum_{g=1}^{s} \epsilon_{igh\alpha\beta} \geq 0, \qquad s = 1, \cdots, N_i - 1$$

while

$$\sum_{g=1}^{N_i} \epsilon_{igh\alpha\beta} = 0.$$

In Theorem 4, we then make the identifications $n = N_i$, $m = N_j$,

$$w_g = \epsilon_{igh\alpha\beta}, \qquad z_h = \epsilon_{jh\alpha\beta} \quad \text{and} \quad a_{gh} = x_{ig} x_{jh}.$$

Then

$$\delta_{gh} = (x_{ig} - x_{i,g+1})(x_{jh} - x_{j,h+1})$$

and hence to have negative correlation between the strata, it is sufficient that the sets $x_{i1}, \cdots, x_{iN_i}$ and $x_{j_1}, \cdots, x_{jN_j}$ have the type of negative ordering represented by $\delta_{gh} \leq 0$. Similarly, if

$$\sigma_{gh} = \mathfrak{E}(x_{ig} - \mu_i)(x_{jh} - \mu_j), \qquad \mu_i = \mathfrak{E}x_{ig},$$

then, for expected negative correlation, it is sufficient that

$$\sigma_{gh} - \sigma_{g,h+1} - \sigma_{g+1,h} + \sigma_{g+1,h+1} \leq 0.$$

Of course, these conditions will be satisfied if a concave upwards correlogram exists. Hence, if a population consists of N random variables $x_1, \cdots, x_N$ having a concave upwards correlogram, then, no matter into what strata these elements are classified, provided that the order of occurrence of the elements remains unaltered, the systematic selection of the elements in the sample can be so planned as to yield an estimate having smaller variance than the stratified random selection of the elements in the sample even if optimum allocation is used. If more than one element is being selected from a stratum under optimum allocation, then the systematic selection of the same number of elements will suffice. If not only optimum allocation but also optimum definitions of strata are being used so that but one element is selected from each stratum, then systematic selection according to the scheme described in the section will produce a variance not larger than the variance of stratified random sampling. It should be noted, however, that this does not imply that a 'hammer and tongs' use of systematic sampling ignoring the strata will produce a smaller variance. There is work to be done on what is required for the latter to occur.

It may be noted that the procedure of this example provides an answer to the systematic selection of elements from a population whose size is not a multiple of the size of sample.

5. The systematic sampling of clusters with probability proportionate to a measure of size. It is known [5] that sampling clusters with probability proportionate to a measure of size often yields considerable reductions in the variance of the estimates. However, the theory of the systematic selection of several clusters with probability proportionate to a measure of size has not been worked out, and it is the purpose of this section to make some contributions to that theory.

The most frequently used method of sampling clusters with probability proportionate to size is equivalent to the following. Suppose that the clusters are denoted by $C_1, \cdots, C_M$ and that to the hth of these M clusters is assigned a measure of size P_h. Form the successive totals $P_1, P_1 + P_2, P_1 + P_2 + P_3, \cdots, P_1 + \cdots + P_M$. If we wish to select m of these clusters, we calculate $\overline{P}_m = (P_1 + \cdots + P_M)/m$. Then, assuming that $P_j \leq \overline{P}_m, j = 1, \cdots, M$, we select an integer with equal probability from $1, \cdots, \overline{P}_m$. Calling that integer P', we calculate the m numbers $P', P' + \overline{P}_m, P' + 2\overline{P}_m, \cdots, P' + (m-1)\overline{P}_m$ If

$$(5.1) \qquad P_1 + \cdots + P_{h-1} + 1 \leq P' + (i-1)\overline{P}_m \leq P_1 + \cdots + P_h$$

for any integer i, $i = 1, \cdots, m$, then the cluster C_h is selected for the sample. Any cluster for which $P_h > \overline{P}_m$ is automatically included in the sample, and if there are, say, α such clusters, then we calculate $\overline{P}_{m-\alpha}$ for the $M - \alpha$ clusters remaining after including these α in the sample, and proceed as above.

In deriving the variance of the estimate we shall use, we interpret that estimate as a stratified sampling estimate Although it is easy to obtain the expected value of the estimate without that interpretation, we shall need it later in the derivation of the variance, and hence we give it here to shorten the total presentation a little.

Suppose that clusters $C_1, \cdots, C_{k_1}$ are such that

$$P_1 + \cdots + P_{k_1-1} < \overline{P}_m \leq P_1 + \cdots + P_{k_1}.$$

Then we define stratum 1 to consist of clusters $C_1, \cdots, C_{k_1}$. It is easy to see that if the above sampling method is used then

$$P\{C_h \text{ is selected from stratum 1}, h < k_1\} = \frac{P_h}{\overline{P}_m},$$

$$P\{C_{k_1} \text{ is selected from stratum 1}\} = \frac{\overline{P}_m - P_1 - \cdots - P_{k_1-1}}{\overline{P}_m}.$$

Furthermore, suppose that clusters $C_{k_1}, \cdots, C_{k_1+k_2}$ are such that

$$P_1 + \cdots + P_{k_1+k_2-1} < 2\overline{P}_m \leq P_1 + \cdots + P_{k_1+k_2}$$

Then we define stratum 2 to consist of clusters $C_{k_1}, \cdots, C_{k_1+k_2}$. It is easy see that if the above sampling method is used, then

$$P\{C_{k_1} \text{ is selected from stratum 2}\} = \frac{P_1 + \cdots + P_{k_1} - \overline{P}_m}{\overline{P}_m},$$

$$P\{C_{k_1+h} \text{ is selected from stratum 2}, 1 \leq h < k_2\} = \frac{P_h}{\overline{P}_m},$$

$$P\{C_{k_1+k_2} \text{ is selected from stratum 2}\} = \frac{2\overline{P}_m - P_1 - \cdots - P_{k_1+k_2-1}}{\overline{P}_m}.$$

Since $P_h \leq \overline{P}_m$ we remark that it is impossible that C_{k_1} be selected from both stratum 1 and stratum 2.

In general, if clusters $C_{k_1+\cdots+k_{i-1}}, \cdots, C_{k_1+\cdots+k_i}$ are such that

$$P_1 + \cdots + P_{k_1+\cdots+k_{i-1}} < i\,\overline{P}_m \leq P_1 + \cdots + P_{k_1+\cdots+k_i} \tag{5.2}$$

then the ith stratum consists of these $k_i + 1$ clusters, and we define the probabilities $P_{i\alpha}$, $\alpha = 0, \cdots, k_i$, by the equations

$$P_{i0} = P\{C_{k_1+\cdots+k_{i-1}} \text{ is selected from stratum } i\} = \frac{P_1 + \cdots + P_{k_1+\cdots+k_{i-1}} - (i-1)\overline{P}_m}{\overline{P}_m},$$

$$\begin{aligned} P_{i\alpha} &= P\{C_h \text{ is selected from stratum } i,\ k_1 + \cdots + k_{i-1} < h < k_1 + \cdots + k_i\} \\ &= \frac{P_h}{\overline{P}_m}, \qquad \alpha = h - k_1 - \cdots - k_{i-1}, \end{aligned} \tag{5.3}$$

$$P_{ik_i} = P\{C_{k_1+\cdots+k_i} \text{ is selected from stratum } i\} = \frac{i\overline{P}_m - P_1 - \cdots - P_{k_1+\cdots+k_{i-1}}}{\overline{P}_m}.$$

We remark that

$$P_{i-1k_{i-1}} + P_{i0} = \frac{P_{k_1+\cdots+k_{i-1}}}{\overline{P}_m}. \tag{5.4}$$

Now, let the elements of the population be x_{hj}, $h = 1, \cdots, M$, $j = 1, \cdots, N_h$, and let the arithmetic mean of the hth cluster be denoted by $\bar{x}_h$. Since the N_h are usually unknown but the measure of size, P_h, is known, we sample, not with probability proportionate to the N_h, but with probability proportionate to the P_h. We shall denote the clusters of the ith stratum by $C_{i0}, \cdots, C_{ik_i}$, making the identification

$$C_{i\alpha} = C_{\alpha+k_1+\cdots+k_{i-1}}. \tag{5.5}$$

Furthermore, the number of elements of the clusters are denoted by $N_{i0}, \cdots,$

N_{ik_i}, and the means of the clusters by $\bar{x}_{i0}, \cdots, \bar{x}_{ik_i}$, where

$$N_{i\alpha} = N_{\alpha+k_1+\cdots+k_{i-1}}$$

$$\bar{x}_{i\alpha} = \bar{x}_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.6}$$

so that $\bar{x}_{i0} = \bar{x}_{i-1,k_{i-1}}$ and $N_{i0} = N_{i-1,k_{i-1}}$, $i = 1, \cdots, m$.

Furthermore, we define

$$\tilde{x}_{i\alpha} = N_{i\alpha}\,\bar{x}_{i\alpha}/P_{i\alpha} = \tilde{x}_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.7}$$

We define the mean of the ith stratum to be

$$\bar{\bar{x}}_i = \sum_{\alpha=0}^{k_i} P_{i\alpha}\tilde{x}_{i\alpha}/\bar{P}_m, \tag{5.8}$$

and the variance of the ith stratum to be

$$\sigma_i^2 = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m}(\tilde{x}_{i\alpha} - \bar{\bar{x}}_i)^2. \tag{5.9}$$

Then, if the mean and variance of the population are defined to be

$$\tilde{x} = \sum_{h=1}^{M} P_h\tilde{x}_h/P \tag{5.10}$$

and

$$\sigma^2 = \sum_{h=1}^{M} \frac{P_h}{P}(\tilde{x}_h - \tilde{x})^2, \tag{5.11}$$

it is easy to verify that

$$\tilde{x} = \frac{1}{m}\sum_{i=1}^{m} \bar{\bar{x}}_i \tag{5.12}$$

and

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 + \frac{1}{m}\sum_{i=1}^{m} (\bar{\bar{x}}_i - \tilde{x})^2. \tag{5.13}$$

An unbiased estimate of the total of a characteristic. We shall see that we can obtain an estimate of x, where

$$x = \sum_{i=1}^{M}\sum_{j=1}^{N_i} x_{ij}$$

i. e. x is the total of the elements of the population. Since N is unknown, the estimate of $\bar{x}$ that is used is the ratio of unbiased estimates of x and N. It is well known that this ratio is usually biased. Since we are not making any study of ratio estimates here we will not derive the approximation to the variance of this estimate. It may be remarked that it can be obtained by a simple extension of the results here given.

Let us agree that the general form of the estimate will be as follows:

If the jth cluster of the population is selected we shall subsample n_j elements from it. The total of the values of the characteristic for these n_j elements we denote by x_j'. Furthermore, we denote by n_i' the total number of elements subsampled from the ith stratum, or, what is the same, from the cluster selected from the ith stratum; and by x_i'' the total of these elements. Thus, if the jth cluster is the ith selected, then $n_i' = n_j$ and $x_i'' = x_j'$. We define our estimate x'' of x, the total of the population, to be

$$x'' = K(x_1'' + \cdots + x_m'') . \tag{5.14}$$

Then, if $K = P/mn$ and $n_h = nN_h/P_h$, it is easy to see that x'' is an unbiased estimate of x.

The variance of the estimate. We may calculate the variance of x'' where

$$x'' = \bar{P}_m (\tilde{x}_1'' + \cdots + \tilde{x}_m'') \quad \text{and} \quad \tilde{x}_i'' = x_i''/n. \tag{5.15}$$

Now, by Theorem 2,

$$\sigma_{x''}^2 = E\sigma_{x''|A}^{2*} + \sigma_{E^*(x''|A)}^2 , \tag{5.16}$$

where A^* has been defined above We shall not evaluate $E\sigma_{x''|A}^{2*}$ since this involves no new problem for subsampling methods using random or systematic methods, or methods using probability proportionate to size.

From (5.15) it follows that

$$E^*(x''|A) = \bar{P}_m(\tilde{x}_1' + \cdots + \tilde{x}_m') \tag{5.17}$$

or, in other words, $E^*(x''|A)$ is the estimate we would have if the clusters in the sample were completely enumerated We shall denote the second term of (5.16) by σ_B^2. Then,

$$\sigma_B^2 = \bar{P}_m^2 \left\{ \sum_{i=1}^{m} \sigma_{\tilde{x}_i'}^2 + \sum_{i \neq j} \sigma_{\tilde{x}_i' \tilde{x}_j'} \right\}. \tag{5.18}$$

Now

$$\sigma_{\tilde{x}_i'}^2 = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m} (\tilde{x}_{i\alpha} - \bar{\bar{x}}_i)^2 = \sigma_i^2. \tag{5.19}$$

To calculate $\sigma_{\tilde{x}_i'\tilde{x}_j'}$, $i \neq j$, we shall use Theorem 1.

$$\sigma_{\tilde{x}_i'\tilde{x}_j'} = E(\tilde{x}_i' - \bar{\bar{x}}_i)(\tilde{x}_j' - \bar{\bar{x}}_j) = E\{(\tilde{x}_i' - \bar{\bar{x}}_i)E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i']\}. \tag{5.20}$$

To calculate $E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i']$ we begin by noting that

$$E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i'] \equiv E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid C_i] \tag{5.21}$$

where C_1^* is the random event having $k_i + 1$ possible states which are the selections of $C_{i0}, \cdots, C_{ik_i}$ as the sample clusters of the ith stratum. Now if $C_{i\alpha}$ is one of the clusters of the ith stratum let us calculate

$$E[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid C_{i\alpha}]. \tag{5.22}$$

We begin by determining which of the clusters of the jth stratum are possible sample clusters, if we know that $C_{i\alpha}$ is selected from the ith stratum. Since the sizes of strata i and j are both $\overline{P}_m$ it follows that there exist integers β_0 and β_1 such that

$$P_{j0} + \cdots + P_{j\beta_0-1} \leq P_{i1} + \cdots + P_{i,\alpha-1} < P_{j1} + \cdots + P_{j\beta_0},$$

and

$$P_{j0} + \cdots + P_{j\beta_1-1} < P_{i1} + \cdots + P_{i\alpha} \leq P_{j1} + \cdots + P_{j\beta_1}.$$

Hence, if we know that $C_{i\alpha}$ has been selected from stratum i, it follows that we must select one of the clusters

$$C_{j\beta_0}, C_{j\beta_0+1}, \cdots, C_{j\beta_1}$$

from stratum j and

$$P\{C_{j\beta} \text{ is selected} \mid C_{i\alpha} \text{ is selected}\} = P'_{j\beta}/P_{i\alpha}, \beta = \beta_0, \beta_0 + 1, \cdots, \beta_1$$
$$= 0, \text{ otherwise,}$$

where

$$P'_{j\beta_0} = P_{j1} + \cdots + P_{j\beta_0} - P_{i1} - \cdots, P_{i,\alpha-1}$$
$$P'_{j\beta} = P_{j\beta}, \beta = \beta_0 + 1, \cdots, B_1 - 1$$
$$P'_{j\beta_1} = P_{i1} + \cdots + P_{i\alpha} - P_{j1} - \cdots - P_{j\beta_1-1},$$

and

$$\sum_{\beta=\beta_0}^{\beta_1} P'_{j\beta} = P_{i\alpha}$$

Then

$$E[(\tilde{x}'_j - \tilde{\bar{x}}_j \mid C_{i\alpha}) = \tilde{x}_{j|\alpha} - \tilde{\bar{x}}_j] \tag{5 23}$$

where

$$\tilde{x}_{j|\alpha} = \sum_{\beta=\beta 0}^{\beta_1} \frac{P'_{j\beta}}{P_{i\alpha}} \tilde{x}_{j\beta}. \tag{5.27}$$

Hence, substituting in (5.20), we see that

$$\sigma_{\tilde{x}'_i\tilde{x}'_j} = E(\tilde{x}'_i - \tilde{\bar{x}}_i)(\tilde{x}'_{j|i} - \tilde{\bar{x}}_j)$$

where $\tilde{x}'_{j|i} = \tilde{x}_{j|\alpha}$ if $C_{i\alpha}$ is selected from stratum i. Then it follows that

$$\sigma_{\tilde{x}'_i\tilde{x}'_j} = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\overline{P}_m} (\tilde{x}_{i\alpha} - \tilde{\bar{x}}_i)(\tilde{x}_{j|\alpha} - \tilde{\bar{x}}_j). \tag{5.25}$$

Obviously, the conditional expectation can be eliminated from (5.25) by using (5 23) but no gain in simplicity or generality thus occurs.

It would be possible to obtain the variances and covariances of the x_i' by listing all possible samples in any special case. To make this general would only require writing the necessary notation.

Substituting in (5.18) we see that

$$\sigma_B^2 = P_m^2 \left\{ \sum_{i=1}^{m} \sigma_i^2 + \sum_{i \neq j} \sigma_{x_i' x_j'} \right\}$$

where $\sigma_{x_i' x_j'}$ is given by (5.25).

It follows that if we use the fact that $\sum_{i=1}^{m} (\bar{\bar{x}}_i - \bar{\bar{x}}) = 0$, then we have

$$\sigma_B^2 = P_m^2 \sum_{i=1}^{m} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P_m} (\bar{x}_{i\alpha} - \bar{\bar{x}})^2 + P_m^2 \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P_m} (\bar{x}_{i\alpha} - \bar{\bar{x}})(\bar{x}_{j|\alpha} - \bar{\bar{x}}),$$

or, returning in part to the "unstratified" notation

$$\sigma_B^2 = \frac{P^2}{m} \sum_{h=1}^{M} \frac{P_h}{P} (\bar{x}_h - \bar{\bar{x}})^2 + \frac{P^2}{m} \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P} (\bar{x}_{i\alpha} - \bar{\bar{x}})(\bar{x}_{j|\alpha} - \bar{\bar{x}}). \tag{5.26}$$

By combining terms of the second part of (4.26) generalizations of the formulae obtained in [1] are easily obtained.

Still another means of writing σ_B^2 is

$$\sigma_B^2 = \frac{P^2}{m} \left\{ \sigma^2 - \sigma_{b \cdot s \cdot}^2 + \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P} (\bar{x}_{i\alpha} - \bar{\bar{x}})(\bar{x}_{j|\alpha} - \bar{\bar{x}}) \right\} \tag{5.27}$$

where

$$\sigma_{b \cdot s \cdot}^2 = \frac{1}{m} \sum_{i=1}^{m} (\bar{\bar{x}}_i - \bar{\bar{x}})^2,$$

which shows both sources of changes in efficiency as compared with sampling with probability proportionate to size, and replacing the clusters obtained. (It is, of course, obvious that $P^2\sigma^2/m$ is the variance of $E^*(x'' \mid A)$, if we assume the m clusters to have been selected with probability proportionate to size, each selected cluster being replaced before the next is selected.)

By considering (5.26) and (5.27) it is clear that systematic sampling with probability proportionate to size will be more efficient than sampling p.p.s. with replacement under much the same conditions as when we sample single elements. The details are omitted. They depend on applying the Lemma and Theorem 4. The summary of the conditions is: If we sample systematically with p.p.s., and if the two sets $x_{i1}, \cdots, x_{ik_i}$ and $y_{j1}, \cdots y_{jk_j}$ are monotone, one being monotone non-icreasing and the other monotone non-decreasing, then the covariance between the ith and jth strata will be negative, and thus gains made as compared with independent sampling from the strata.

If we define

$$\sigma_{\alpha\beta}^0 = \mathcal{E}(\bar{x}_{i\alpha} - \mathcal{E}\bar{x}_{i\alpha})(\bar{x}_{j\beta} - \mathcal{E}\bar{x}_{j\beta})$$

then the concavity condition for systematic sampling p.p.s. to yield a smaller variance than independent sampling p.p.s. from each stratum is, if $\alpha < \beta$,

$$\sigma^0_{\alpha 1} - \sigma^0_{\gamma 1} \leq \sigma^0_{\alpha 2} - \sigma^0_{\gamma 2} \leq \cdots \leq \sigma^0_{\alpha k_f} - \sigma^0_{\gamma k_f} \leq 0.$$

6. The systematic sampling of clusters of equal sizes. Let us now suppose that our population consists of clusters of elements, the clusters being of equal size, i.e. containing the same number of elements. To be specific, let the population consist of M clusters, where $M = cm$ and each cluster contains N elements, where $N = kn$. Then, the value of the characteristic being measured for the αth element of the ith cluster may be denoted by $x_{i\alpha}$, and the total of all the elements of the ith cluster may be denoted by $x_{i\cdot}$. The arithmetic mean of the population is $\bar{x}$, and thus

$$M\bar{x} = \sum_{i=1}^{M} \bar{x}_{i\cdot}.$$

where

$$N_i\bar{x}_{i\cdot} = x_{i\cdot}.$$

a. *Complete enumeration of clusters in sample.* First, suppose that we wish to estimate $\bar{x}$ by $\bar{x}'$, where $\bar{x}'$ is the arithmetic mean of the sample obtained by selecting a systematic sample of m of the M clusters, and enumerating all elements within each cluster in the sample. Then, we may write

$$m\bar{x}' = \sum_{i=1}^{m} \bar{x}'_{i\cdot}, \tag{6.1}$$

where $\bar{x}'_{i\cdot}$ is the mean of the ith cluster selected for the sample. From [1], it follows then that

$$\sigma^2_{\bar{x}'} = \frac{\sigma^2_b}{m}\{1 + (m-1)\bar{\rho}_c\}$$

where $M\sigma^2_b = \sum_{i=1}^{M} (\bar{x}_{i\cdot} - \bar{x})^2$, and $\bar{\rho}_c$ is defined as $\bar{\rho}_k$ in [1, p. 6], but with $\bar{x}_{i\cdot}$ in place of $x_{i\cdot}$. Now from the theory of the random sampling of clusters it follows that

$$\sigma^2_b = \frac{\sigma^2}{N}\{1 + (N-1)\rho\}$$

where σ^2 is the variance of the population, i. e.

$$MN\sigma^2 = \sum_{i=1}^{M}\sum_{j=1}^{N} (x_{ij} - \bar{x})^2$$

and ρ is the intraclass correlation coefficient of elements within clusters, i. e.

$$\sigma^2\rho = \sigma^2_b - \sigma^2_w/N - 1,$$

where

$$MN\sigma_w^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)^2.$$

Thus

$$\sigma_{\bar{x}'}^2 = \frac{\sigma^2}{mN}\{1 + (N-1)\rho\}\{1 + (m-1)\bar{\rho}_c\}. \tag{6.2}$$

Of the three factors in (6.2), σ^2/mN is the variance of a random sample of size mN selected with replacement; $1 + (N-1)\rho$ is the factor arising from the use of clusters; and $1 + (m-1)\bar{\rho}_c$ is the factor arising from the fact that the clusters are sampled systematically.

b. *Stratification and subsampling.* When we consider the possibilities of stratification and subsampling, the number of possible designs increases tremendously. For example, it would be simple to calculate the variances of arithmetic means obtained by stratifying the population, selecting sampling units with probability proportionate to size, subsampling systematically, again subsampling systematically and finally subsampling at random. However, such studies may be left to be made in connection with the practical problems in which they are to be used. Rather than attempt to consider many of the possibilities that might arise in practice, we shall here give only the results of the systematic subsampling of a systematic sample. The variances of many other designs may easily be obtained by means of Theorems 1 and 2.

Suppose now that from each of a systematically selected sample of m clusters we subsample, systematically, n elements. Then, let our estimate of $\bar{x}$ be $\bar{x}'$ where, if $x'_{i\alpha}$ is the αth selected element from the ith sample cluster, then

$$\bar{x}' = \left(\frac{1}{mn}\right)\sum_{i=1}^{m}\sum_{\alpha=1}^{n}x'_{i\alpha} = \frac{1}{m}\sum_{i=1}^{m}\bar{x}''_i$$

and

$$\bar{x}''_i = \left(\frac{1}{n}\right)\sum_{\alpha=1}^{m}x'_{i\alpha}.$$

From Theorem 2, it follows at once that

$$\sigma_{\bar{x}'}^2 = \frac{\sigma^2}{mN}\{1 + (N-1)\rho\}\{1 + (m-1)\bar{\rho}_c\} + \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i^2}{mn}\{1 + (m-1)\bar{\rho}_i\}, \tag{6.3}$$

where σ_i^2 is the variance within the ith cluster and ρ_i is the average serial correlation within the ith cluster as defined in [1, p. 6]. It is simple to calculate the variance of $\bar{x}'$ also when the sub-sampling is done by considering the m clusters in the sample as one population from which a systematic sample is selected. This is the case that occurs when a sample of blocks is selected and all the households on the sample blocks are listed serially, a systematic sample then being selected from the lists. However, for our present purposes it is the analysis of (6.2) that is important and we now turn to a brief discussion of (6 2).

The most important conclusion to be drawn from (6 2) is that the systematic

selection of clusters even when systematic selection is desirable, may not compensate for the increase in variance caused by the use of clusters. Systematic selection will provide the same relative gains but these gains may not be large enough to produce the inequality

$$\{1 + (N - 1)\rho\}\{1 + (m - 1)\bar{p}_c\} < \frac{MN - mN}{MN - 1}.$$

A problem that we have not worked through is the following: By regarding the elements of the population as random variables, we obtain conditions on the average correlations among elements of a single cluster as well as on the average correlations among elements of different clusters that enable us to state where the systematic sampling of clusters of equal sizes may be expected to yield a smaller variance than the random or stratified random sampling of clusters or of individual elements. This solution should be straight forward.

c. *Systematic sampling in two dimensions.* Systematic sampling in two dimensions occurs in such practical problems as the selection of a sample of blocks from a city or the selection of a sample of plots from a field

In selecting blocks from a city, the procedure most often followed effectively reduces the problem to one dimensional form by first numbering the blocks of the city or a part of it, in serpentine fashion beginning, say, in the upper right corner of a map of the city and numbering the blocks in the top row from right to left continuing the numbering of the second row from left to right and so on. Then a systematic sample of these block numbers, and hence, of the blocks themselves is selected. Clearly, this procedure should not be the most efficient if neighboring blocks are highly correlated, since, to cite an unrealistic possibility, the possible samples might turn out to be columns of blocks of the city.

A second two dimensional systematic sampling procedure might be that of selecting a systematic sample of the rows and a systematic sample of the columns, thus obtaining a grid sample This design too is inefficient when there is a "fertility gradient" along rows or along columns.

The reason for the inefficiency of both of these procedures can be found by examining the formulae for the variances of systematic samples. If the numbering is serpentine, then it becomes illogical to expect that the correlogram is concave upwards and sharp deviations from that pattern may occur. In the grid design, which is a special case of the systematic sampling of clusters with systematic subsampling, we may examine (6.3) and note that the intra-class correlation coefficient ρ may be large enough for $\sigma_{\bar{x}}^2$, to be large even when $\bar{p}_c$ is negative.

Clearly, (6 3) suggests that the possible samples be so defined that ρ is as small as possible In square fields this might be attained by defining the possible samples to be plots of a Knut Vik square having the same treatment, and similar definitions of possible samples could easily be given for irregular fields. This subject is, however, left for further study.[3]

[3] One of the referees of this paper has drawn the author's attention to an article [6], the data of which, especially Table 3, are in accordance with the opinions expressed above.

REFERENCES

[1] W. G. MADOW AND L. H. MADOW, "On the theory of systematic sampling, I," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 1–24.

[2] L. H. MADOW, "Systematic sampling and its relation to other sampling designs," *Am. Stat. Asso. Jour.*, Vol. 41 (1946), pp. 204–217.

[3] W G. COCHRAN, "Relative accuracy of systematic and stratified random samples for a certain class of populations," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 164–177.

[4] G. H. HARDY, J. E. LITTLEWOOD AND G. POLYA, *Inequalities*, Cambridge University Press, London and New York, 1934, p. 43.

[5] M. H. HANSEN AND W. N. HURWITZ, "On the theory of sampling from finite populations," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 333–362.

[6] P. G. HOMEYER AND C. A. BLACK, "Sampling replicated field experiments on oats for yield determinations," *Socl. Sci. Soc. Proc.*, Vol. 11 (1946), pp. 341–344.

PROBLEMS IN PLANE SAMPLING

By M. H. Quenouille

Rothamsted Experimental Station, Harpenden, England

1. Summary. After consideration of the relative accuracies of systematic and stratified random sampling in one dimension the problem of estimation of linear sampling error is discussed.

Methods of sampling an area are proposed, and expressions for the accuracies of these methods are derived. These expressions are compared for large samples, with special reference to correlation functions which appear to be theoretically and practically justified, and systematic sampling is found to be more accurate than stratified random sampling in many cases. Methods of estimating sampling errors are again considered, and examples given. The paper concludes with some remarks on the problem of trend in the population sampled.

2. Accuracy of systematic and stratified random samples in one dimension. W. G. Cochran [1] has given expressions to the variances of the means of samples of size n drawn from a population $x_1x_2 \cdots x_{nk}$ when the method of sampling is random (r), stratified random (st) and systematic (sy). He assumes the elements $x_1x_2 \cdots x_{nk}$ to be drawn from a population in which

$$E(x_i) = \mu, \qquad E(x_i - \mu)^2 = \sigma^2, \qquad E(x_i - \mu)\,(x_{i+u} - \mu) = \rho_u\sigma^2$$

where $\rho_u \geq \rho_v \geq 0$ whenever $u < v$, and derives the expressions

$$(1) \qquad \sigma_r^2 = \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right)\left[1 - \frac{2}{kn(kn-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u\right]$$

$$(2) \qquad \sigma_{st}^2 = \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right)\left[1 - \frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)\rho_u\right]$$

$$(3) \qquad \sigma_{sy}^2 = \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right) \cdot \left[1 - \frac{2}{kn(k-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u + \frac{2k}{n(k-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right].$$

Using these expressions which are linear functions of the ρ_u Cochran compares the relative efficiencies of the methods of sampling for several types of correlogram. It is worth noting that (1), (2) and (3) can be derived under more general conditions than Cochran considered. If we assume that (a) each x_i is a sample from a population with mean μ_i and variance σ_i^2, (b) that μ_i is distributed about mean μ with variance σ^2, (c) that $E(\mu_i - \mu)\,(\mu_j - \mu) = \rho_{ij}\sigma^2$, and (d) that $\rho_u = \dfrac{1}{kn-u}\sum_{i=1}^{kn-u}\rho_{i,i+u}$, then it is not difficult to show that (1), (2) and (3)

require the addition of a superposed variation $\frac{1}{n}\left(1-\frac{1}{k}\right)\cdot\frac{1}{kn}\sum_{i=1}^{kn}\sigma_i^2$ to the right-hand side of the equations. Thus it should be remembered that Cochran's results give theoretical maxima to the relative efficiencies of the various methods of sampling, while ρ_u is the mean correlation between samples u apart. This result is perhaps interesting in connection with sampling for say, insect infestation, when at each point there will be a mean level of infestation and the sample will be distributed in a Poisson distribution about this mean. Then the superposed variation is

$$\frac{1}{n}\left(1-\frac{1}{k}\right)\cdot\frac{1}{kn}\sum_{i=1}^{kn}\mu_i\sim\frac{1}{n}\left(1-\frac{1}{k}\right)\mu.$$

If we are sampling a continuous process[1], for n large we can write down the integral equivalents of (1), (2) and (3)

$$\sigma_r^2\sim\frac{\sigma^2}{n}$$

$$\sigma_{st}^2\sim\frac{\sigma^2}{n}\left[1-\frac{2}{d^2}\int_0^d(d-u)\rho_u\,\delta u\right] \tag{4}$$

$$\sigma_{su}^2\sim\frac{\sigma^2}{n}\left[1-\frac{2}{d}\int_0^\infty\rho_u\,\delta u+2\sum_{u=1}^{\infty}\rho_{du}\right] \tag{5}$$

where ρ_u is the mean correlation between successive elements of the sample, u apart and d is the mean distance between samples. We have thus

$$\frac{\sigma_{st}^2-\sigma_{su}^2}{\sigma_r^2}\sim\frac{2}{d}\left[\int_0^d\frac{u}{d}\rho_u\,\delta u+\int_d^\infty\rho_u\,\delta u-d\sum_{u=1}^{\infty}\rho_{du}\right],$$

which can often be used to investigate, quickly and roughly, with the aid of a graph the difference between the efficiencies of stratified random and systematic sampling. Figure 1 shows how this is done for four types of correlogram.

For a continuous Markoff scheme, we have $\rho_u=\rho^u$ and

$$\sigma_{st}^2\sim\frac{\sigma^2}{n}\left[1+\frac{2}{\log\rho^d}+\frac{2}{(\log\rho^d)^2}-\frac{2\rho^d}{(\log\rho^d)^2}\right],$$

$$\sigma_{su}^2\sim\frac{\sigma^2}{n}\left[1+\frac{2}{\log\rho^d}+\frac{2\rho^d}{1-\rho^d}\right],$$

which agree with Cochran's results.

3. Replication and the estimation of error. Yates [2] has pointed out the difficulties attached to the estimation of error for a systematic sample. It will, however, be worthwhile to investigate this point using the above formulae.

[1] In practice we can sample a continuous process only as if it were a discontinuous process with k large.

For random, stratified random and systematic sampling, if n is large and k is regarded as constant, then the variance of the estimate of the mean will be of the form $\sigma^2 F(k)/n$, where $F(k)$ is virtually independent of n. Thus, if we have any method which provides an estimate of error for the samples it will be possible to split the series to be sampled into several equal parts (or blocks) to obtain an estimate of error of the mean of each part and to combine these to obtain a more accurate estimate of the error of the overall mean. In fact, if n is very large, we may wish to reduce our number of observations by obtaining estimates of error from a random selection of these parts. For stratified random sampling, $F(k)$ is completely independent of n, so that we may combine our estimates of error from each strata. This leads us to the commonly used method of taking q randomly chosen elements per strata, and combining the sets of variances of $q - 1$ degrees of freedom to form an estimate of error. If we make our samples exclusive, i.e. no two elements can coincide, then this variance has to be multiplied by $1 - q/k$ to give the estimated variance of the sample mean.

We can in the same way estimate the variance of the mean of a systematic sample by using sets of q systematic samples of sufficient length with randomly-chosen starting points. This sampling will, however, be more difficult to carry out in practice, and we might consider other methods. Our systematic samples may be chosen to be invariable in each part or block into which the series is split so that our sampling procedure involves, in all, only q systematic samples, or we might follow the method advocated by Yates of choosing our q samples to be evenly spaced, so that they are subsamples of a larger systematic sample. Whereas this latter method has simplicity and its possible incorporation into a more extensive scheme to recommend it, its use has to be very carefully considered. If we consider the discrete case, we wish to estimate

$$\sigma^2 \left(1 - \frac{2}{k-1} \sum_{u=1}^{\infty} \rho_u + \frac{2k}{k-1} \sum_{u=1}^{\infty} \rho_{ku}\right), \tag{6}$$

but any estimate of variance based on q evenly-spaced systematic samples can contain only terms of the form $\rho_{ku/q}$, and while an estimate of variance based on q randomly-chosen systematic samples will obviously be limited, it will, in most cases, be more representative. As an example, suppose we take $k = 16$ and $q = 4$ then we can compare the relative occurrences of observing the correlations $\rho_1 \cdots \rho_{15}$ in the estimate of variance. Six examples of this are given in table 1, the random numbers having been drawn from Fisher and Yates tables, ρ_u and ρ_{16-u} being shown together, since they occur equally frequently. The table demonstrates how randomly-chosen samples, even as nearly systematic as the first two randomly-chosen samples will avoid systematically sampling the correlogram. It is obvious that in most cases either method will be fairly good but the use of this latter will usually be the more accurate. Comparisons are made in table 2 for various types of correlogram using the samples indicated in table 1. It is, of course, possible to postulate theoretically many kinds of

[2] Throughout this paper δ is used for the differential sign to prevent confusion with d.

correlogram for which the equal-spaced sets of systematic samples will break down, but ultimately we must decide with reference to the types of correlogram

TABLE 1

Frequency of occurrence of the serial correlations ρ_1, $\rho_2 \ldots \rho_{15}$ *in the estimate of variance when 4 systematic samples each with spacing 16 units are taken*

ρ	4 evenly-spaced systematic samples	4 systematic samples with random starting points at						Total frequencies
		4, 7, 8, 12,	3, 7, 8, 12,	3, 6, 10, 13,	4, 6, 7, 14,	2, 8, 11, 15,	2, 6, 11, 16	
1, 15		1	1		1			3
2, 14					1		1	2
3, 13		1		2	1	2		6
4, 12	4	2	2	1		1	1	7
5, 11		1	2				2	5
6, 10				1	1	1	1	4
7, 9			1	2	1	2	1	7
8	4	2			2			4

TABLE 2

Values of $\frac{1}{15}\sum_{u=1}^{15} \rho_u$ *as estimated by systematic samples*

ρ_u	Evenly-spaced systematic samples	Systematic samples with random starting points						Mean	Expected
		1	2	3	4	5	6		
$1-0.2\,u$, $(u = 1, .5)$	0.17	0.27	0 20	0.17	0.30	0.17	0.13	0.21	0.27
$1-0.1\,u$, $(u = 1, .10)$	0.53	0.62	0.58	0.53	0.60	0.53	0.53	0.57	0.60
2^{-u}	0.04	0.13	0.12	0.06	0.15	0.06	0.07	0.10	0.13
$2^{-u/4}$	0.58	0.66	0.64	0.60	0.66	0.60	0.60	0.63	0.65
Kendall's Series 1	−0.14	0.03	0.00	−0.05	0.16	−0.05	−0.05	0.01	0.07

* Naturally the use of this method of estimating the sampling error assumes that the correlation between the corresponding elements in each part or block into which the series is split may be neglected, i.e. in this case that the terms ρ_{16} and above are negligible. In this case $\rho_{16} = 1/16$ and consequently the term $2(\frac{1}{15}\sum_{u=1}^{\infty} \rho_u - \frac{1}{15}\sum_{u=1}^{\infty} \rho_{16u}) = 0.56$, required in (6) differs slightly from the term $\frac{1}{15}\sum_{u=1}^{15} \rho_u = 0.65$ which we are attempting to estimate.

experienced. We shall consider this point further, after we have dealt with two-dimensional sampling.

4. Methods of sampling in 2 dimensions. The number of ways in which we can sample a two-dimensional space[3] is large, since we can employ random,

[3] We shall, in general, consider our two-dimensional space to be rectangular, but it is not difficult to draw similar conclusions for an area of any shape.

stratified random or systematic sampling in either direction. Thus we will be able to consider every possible combination of these methods, e.g. random in

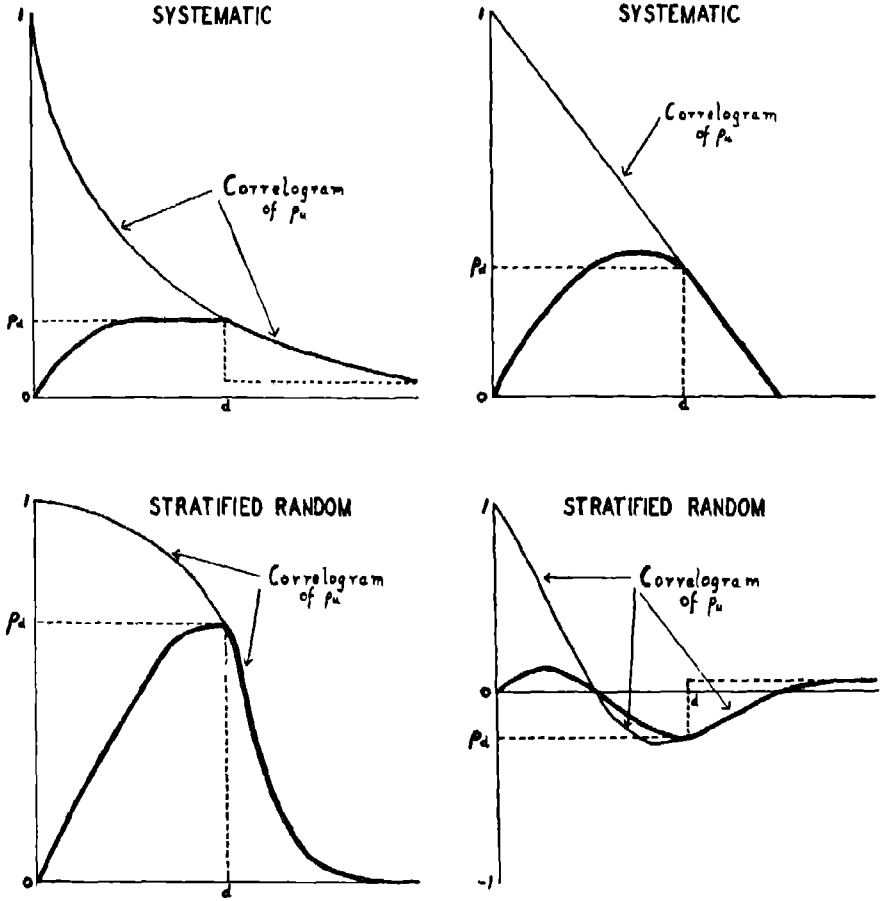

FIG. 1. Graphical comparison of the efficiencies of systematic and stratified random sampling for various correlation functions. The thick line gives the function

$$f_1(u) = u\rho_u/d, \qquad 0 \leq u \leq d$$
$$= \rho_u, \qquad d \leq u,$$

and the dotted line the function

$$f_2(u) = \rho_{id}, \qquad (i-1)d < u \leq id.$$

Thus systematic sampling is more or less efficient than stratified random sampling according to whether the area under the thick line is greater or less than the area under the dotted line. The most efficient method is indicated on each graph.

one direction and systematic in another will be denoted by $r \cdot sy$. Furthermore we can consider the sets of samples in one direction to be aligned with one another, or to be independently determined. The suffix 1 will be used to denote

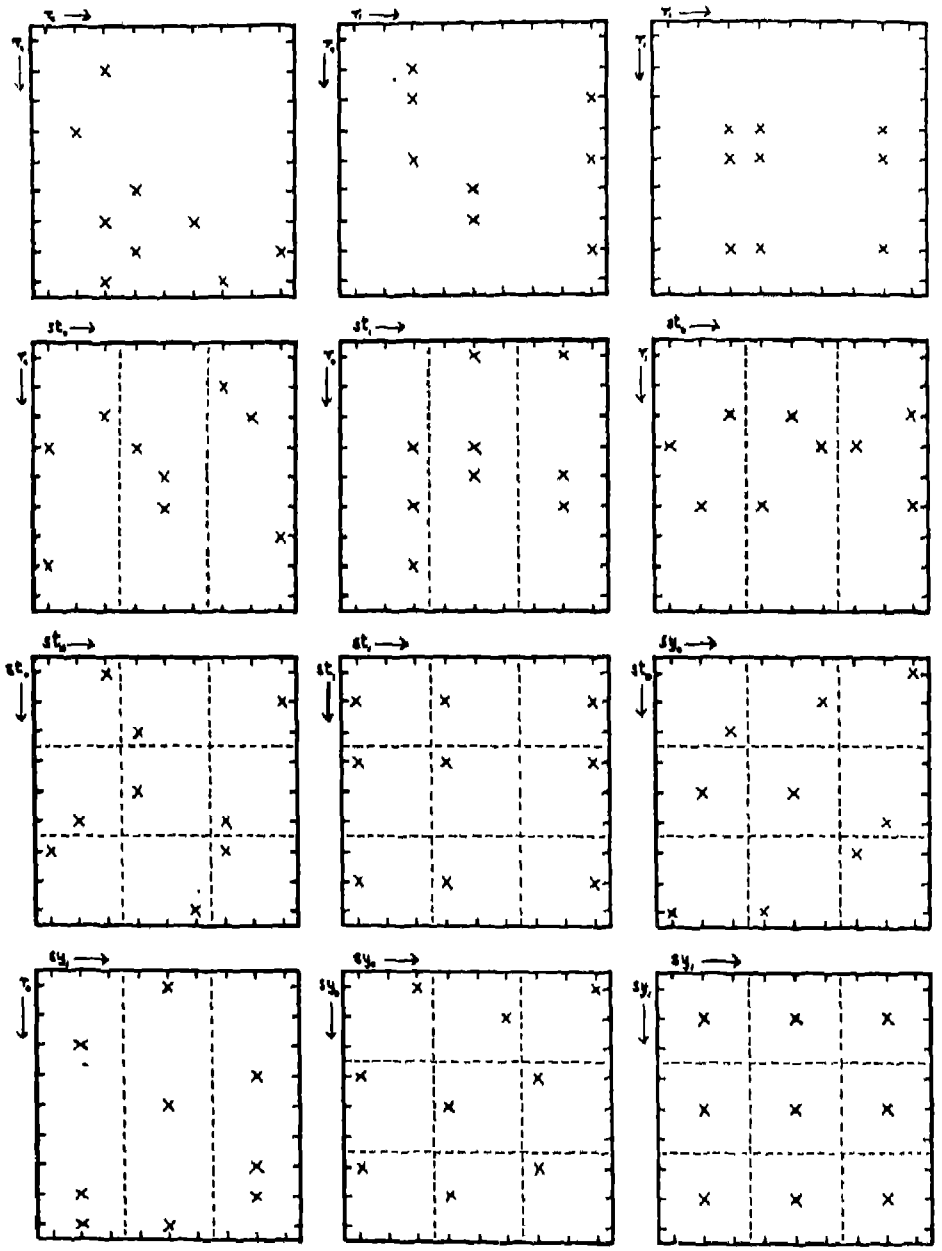

FIG. 2. Methods of sampling a field. In this case, $n_1 = n_2 = k_1 = k_2 = 3$.

aligned samples while suffix 0 will denote independent samples, e.g. we might sample according to the system $r_1 s y_0$. Examples of several methods of sampling are given in Figure 2.

5. Accuracy of sampling in two dimensions. Suppose we consider a sample of n_1n_2 elements drawn from the elements $x_{i,j}(i = 1, 2, \cdots n_1k_1, j = 1, 2, \cdots n_2k)$, (which form a single finite population drawn from an infinite hypothetical population), such that the mean spacing in the two directions is k_1 and k_2. These parameters will, if necessary, be indicated in brackets after the method of sampling, e g. $r_1sy_0(n_1k_1, n_2k_2)$.

Let X denote the mean of a sample formed by the method considered, and x' a member of this sample Suppose, also, that the $x_{i,j}$ are drawn from a population in which

$$E(x_{i,j}) = \mu, \qquad E(x_{i,j} - \mu)^2 = \sigma^2,$$
$$E(x_{i,j} - \mu)\,(x_{i+u,j+v} - \mu) = \rho_{i,juv}\sigma^2,$$

Further we may average $\rho_{i,juv}$ over all possible values of i and j to define $\rho_{uv} = \rho_{-u,-v}$ by the relation

$$\sum_i \sum_j \rho_{ijuv} = (k_1n_1 - |u|)(k_2n_2 - |v|)\rho_{uv}.$$

The purpose of these definitions is to allow to eliminate the difficulties associated with the parameters of finite populations by considering this population as being itself a sample from an infinite population. Cochran employs a similar device.

5a. *Random sampling.* It is not difficult to see that

$$\sigma^2(X) = \tfrac{1}{2}E(X_1 - X_2)^2 = E(X_1 - \mu)^2 - E(X_1 - \mu)\,(X_2 - \mu),$$

where X_1 and X_2 are independent samples.

Also

$$E(X_1 - \mu)(X_2 - \mu) = E(x_1' - \mu)(x_2' - \mu)$$
$$= \frac{\sigma^2}{k_1k_2n_1n_2}\left[1 + \frac{1}{k_1k_2n_1n_2}\sum\sum (k_1n_1 - |u|)(k_2n_2 - |v|)\rho_{uv}\right]$$

where the double summation[4] exists over the region S given by $|u| \leq k_1n_1$, $|v| \leq k_2n_2$ and excludes $u = v = 0$. We thus have to evaluate $E(X_1 - \mu)^2$ for the different types of random sampling.

It is easily shown that

$$E(X_1 - \mu)^2 = \frac{\sigma^2}{n_1n_2}$$
$$\cdot\left[1 + \frac{n_1n_2 - 1}{k_1k_2n_1n_2(k_1k_2n_1n_2 - 1)}\sum\sum (k_1n_1 - |u|)(k_2n_2 - |v|)\rho_{uv}\right]$$
for r_0r_v,
$$= \frac{\sigma^2}{n_1n_2}\left[1 + \frac{n_1 - 1}{k_1k_2n_1(k_1k_2n_1n_2 - 1)}\sum\sum (k_1n_1 - |u|)(k_2n_2 - |v|)\rho_{uv}\right.$$
$$\left. + \frac{2(n_2 - 1)}{k_2n_2(k_2n_2 - 1)}\sum_{v=1}^{k_2n_2}(k_2n_2 - v)\rho_{0v}\right]$$

[4] In general, unless otherwise stated, double summations will exist over the region for which the coefficients are positive, excluding $u = v = 0$

for $r_1 r_0$,

$$= \frac{\sigma^2}{n_1 n_2}\Big[1 + \frac{(n_1-1)(n_2-1)}{k_1 k_2 n_1 n_2(k_1 k_2 n_1 n_2 - 1)} \sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}$$

$$+ \frac{2(n_2-1)}{k_2 n_2(k_2 n_2 - 1)} \sum_{v=1}^{k_2 n_2} (k_2 n_2 - v)\rho_{0v} + \frac{2(n_1-1)}{k_1 n_1(k_1 n_1 - 1)} \sum_{u=1}^{k_1 n_1} (k_1 n_1 - u)\rho_{u0}\Big]$$

for $r_1 r_1$,

whence

$$\sigma^2(r_0 r_0) = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2$$

$$(7) \qquad \cdot\left[1 - \frac{1}{k_1 k_2 n_1 n_2(k_1 k_2 n_1 n_2 - 1)} \sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}\right]$$

$$\sigma^2(r_1 r_0) = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\Big[1 - \frac{k_1 k_2 n_2 - 1}{(k_1 k_2 - 1)k_1 k_2 n_1 n_2(k_1 k_2 n_1 n_2 - 1)}$$

$$(8) \qquad \cdot\sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}$$

$$+ \frac{2(n_2-1)}{k_2 n_2(k_2 n_2 - 1)} \sum_{1}^{k_2 n_2} (k_2 n_2 - v)\rho_{0v}\Big]$$

$$\sigma^2(r_1 r_1) = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\Big[1 - \frac{k_1 k_2(n_1 + n_2 - 1) - 1}{(k_1 k_2 - 1)k_1 k_2 n_1 n_2(k_1 k_2 n_1 n_2 - 1)}$$

$$(9) \qquad \cdot\sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv} + \frac{2k_1(n_2-1)}{(k_1 k_2 - 1)n_2(k_2 n_2 - 1)}$$

$$\cdot\sum_{v=1}^{k_2 n_2} (k_2 n_2 - v)\rho_{0v} + \frac{2k_2(n_1-1)}{(k_1 k_2 - 1)n_1(k_1 n_1 - 1)} \sum_{u=1}^{k_1 n_1} (k_1 n_1 - u)\rho_{u0}\Big].$$

5b. *Stratified random sampling.* We can deduce the variances for some methods of taking stratified samples if $\bar{x}_i'$, the mean of the elements sampled in the ith stratum, is independent of $\bar{x}_j'$, since we will then have

$$E(X - \bar{x})^2 = E(\bar{x}_i' - \bar{x}_i)^2/n,$$

where $\bar{x}$ is the mean of the finite population which is sampled. Hence

$$\sigma^2(st_0 r_0) = \frac{1}{n_1}\sigma^2\{r_0 r_0(1, k_1 ; n_2 k_2)\}$$

$$(10) \qquad = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\Big[1 - \frac{1}{k_1 k_2 n_2(k_1 k_2 n_2 - 1)}$$

$$\cdot\sum\sum (k_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}\Big],$$

$$\sigma^2(st_1\, r_0) = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\left[1 - \frac{1}{k_1 k_2 n_2(k_1 k_2 - 1)}\right.$$

$$\cdot \sum\sum (k_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}$$

$$\left. + \frac{2k_1(n_2 - 1)}{(k_1 k_2 - 1)n_2(k_2 n_2 - 1)} \sum_{v=1}^{k_2 n_2} (k_2 n_2 - v)\rho_{uv}\right], \tag{11}$$

$$\sigma^2(st_0\, st_0) = \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\left[1 - \frac{1}{k_1 k_2(k_1 k_2 - 1)}\right.$$

$$\left. \cdot \sum\sum (k_1 - |u|)(k_2 - |v|)\rho_{uv}\right]. \tag{12}$$

To estimate the variance of other methods of sampling, we will make use of a general formula which we might have used to derive the expressions (8)–(12).

If x_i' is any element of the sample X, then

$$(X - \bar{x})^2 = \frac{1}{n_1 n_2}\left[\sum (x_i' - \bar{x})^2 - \sum (x_i' - X)^2\right]$$

$$= \frac{1}{n_1 n_2}\left[\sum (x_i' - \bar{x})^2 - \frac{n_1 n_2 - 1}{n_1 n_2}\sum (x_i' - \mu)^2\right.$$

$$\left. + \frac{2}{n_1 n_2}\sum\sum (x_i' - \mu)(x_j' - \mu)^3\right],$$

whence

$$\sigma^2(X) = E(X - \bar{x})^2$$

$$= \frac{k_1 k_2 n_1 n_2 - 1}{k_1 k_2 n_1 n_2}\sigma^2\left[1 - \frac{1}{k_1 k_2 n_1 n_2(k_1 k_2 n_1 n_2 - 1)}\sum\sum (k_1 n_1 - |u|)\right.$$

$$\left. \cdot (k_2 n_2 - |v|)\rho_{uv}\right] - \frac{n_1 n_2 - 1}{n_1 n_2}\sigma^2 + \frac{n_1 n_2 - 1}{n_1 n_2} E(x_i' - \mu)(x_j' - \mu) \tag{13}$$

$$= \frac{1}{n_1 n_2}\left(1 - \frac{1}{k_1 k_2}\right)\sigma^2\left[1 - \frac{1}{k_1 k_2 n_1 n_2(k_1 k_2 - 1)}\sum\sum (k_1 n_1 - |u|)\right.$$

$$\left. \cdot (k_2 n_2 - |v|)\rho_{uv} + \frac{k_1 k_2(n_1 n_2 - 1)}{k_1 k_2 - 1}\frac{E(x_i' - \mu)(x_j' - \mu)}{\sigma^2}\right].$$

Thus, provided that we can estimate $E(x_i' - \mu)\,(x_j' - \mu)/\sigma^2$ the expression (13) gives the error for all methods of sampling.

As an example, we might deduce the expression (12). If we choose any member x_i', then a second member x_j' will be located at random with respect to x_i' except that there will be $k_1 k_2 - 1$ positions in the same stratum as x_i' that x_j' will not be able to occupy. Thus the expected correlation $E(x_i' - \mu)\,(x_j' - \mu)/\sigma^2$ will be given by

$$\frac{1}{k_1^2 k_2^2 n_1 n_2(n_1 n_2 - 1)}\sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}$$

$$- \frac{1}{k_1^2 k_2^2(n_1 n_2 - 1)}\sum\sum (k_1 - |u|)(k_2 - |v|)\rho_{uv}. \tag{14}$$

If we substitute (14) into (13), we will obtain expression (12) for the variance of st_0st_0. In the same manner, we can derive for st_1st_1 the expression

$$
\begin{aligned}
&E(x_i' - \mu)(x_j' - \mu) \\
&= \frac{1}{k_1 k_2(n_1 n_2 - 1)}\left[\frac{1}{k_1 k_2 n_1 n_2}\sum\sum (k_1 n_1 - |u|)(k_2 n_2 - |v|)\rho_{uv}\right. \\
&\qquad - \frac{1}{k_1 k_2 n_1}\sum\sum (k_1 n_1 - |u|)(k_2 - |v|)\rho_{uv} \\
&\qquad - \frac{1}{k_1 k_2 n_2}\sum\sum (k_1 - |u|)(k_2 n_2 - |v|)\rho_{uv} + \frac{1}{k_1 k_2}\sum\sum (k_1 - |u|) \\
&\qquad \cdot (k_2 - |v|)\rho_{uv} + \frac{2(k_1 k_2 n_1 - 1)}{k_1 n_1(k_1 n_1 - 1)}\sum_{u=1}^{k_1 n_1}(k_1 n_1 - u)\rho_{0v} \\
&\qquad - \frac{2(k_1 k_2 - 1)}{k_1(k_1 - 1)}\sum_{u=1}^{k_1}(k_1 - u)\rho_{uv} + \frac{2(k_1 k_2 n_2 - 1)}{k_2 n_2(k_2 n_2 - 1)}\sum_{v=1}^{k_2 n_2}(k_2 n_2 - v)\rho_{0v} \\
&\qquad \left. - \frac{2(k_1 k_2 - 1)}{k_2(k_2 - 1)}\sum_{v=1}^{k_2}(k_2 - v)\rho_{uv}\right].
\end{aligned} \tag{15}
$$

Thus we can evaluate $\sigma^2(X)$ for all types of stratified random sampling

5c. *Systematic sampling.* In a similar manner to that used for stratified random sampling, we can use (13) to evaluate the variances of systematic sampling. Values of $E(x_i' - \mu)(x_j' - \mu)$ for three of the possible methods of sampling are given below. For sy_1sy_1

$$
E(x_i' - \mu)(x_j' - \mu) = \frac{1}{n_1 n_2(n_1 n_2 - 1)}\sum\sum(n_1 - |u|)(n_2 - |v|)\rho_{k_1 u, k_2 v} \tag{16}
$$

For sy_1r_0

$$
\begin{aligned}
&E(x_i' - \mu)(x_j' - \mu) = \frac{1}{k_2^2 n_1 n_2(n_1 n_2 - 1)}\sum\sum (n_1 - |u|) \\
&\cdot (k_2 n_2 - |v|)\rho_{k_1 u, v} - \frac{2(k_2 - 1)}{k_2^2 n_2(n_1 n_2 - 1)(k_2 n_2 - 1)}\sum_{v=1}^{k_2 n_2}(k_2 n_2 - v)\rho_{0v}.
\end{aligned} \tag{17}
$$

For sy_0sy_0

$$
\begin{aligned}
&E(x_i' - \mu)(x_j' - \mu) = \frac{1}{k_1 k_2(n_1 n_2 - 1)}\left[\frac{1}{k_1 k_2 n_1 n_2}\sum\sum (k_1 n_1 - |u|)\right. \\
&\cdot (k_2 n_2 - |v|)\rho_{uv} - \frac{1}{k_1 k_2 n_1}\sum\sum (k_1 n_1 - |u|)(k_2 - |v|)\rho_{uv} \\
&- \frac{1}{k_1 k_2 n_2}\sum\sum (k_1 - |u|)(k_2 n_2 - |v|)\rho_{uv} + \frac{1}{k_1 k_2}\sum\sum (k_1 - |u|) \\
&\cdot (k_2 - |v|)\rho_{uv} + \frac{k_1}{k_2 n_1}\sum\sum (n_1 - |u|)(k_2 - |v|)\rho_{k_1 u, v}
\end{aligned} \tag{18}
$$

$$- \frac{2k_1}{k_2} \sum_{v=1}^{k_2} (k_2 - v)\rho_{0v} + \frac{k_2}{k_1 n_2} \sum\sum (k_1 - |u|)(n_2 - |v|)\rho_{u_1 k_2 v} - \frac{2k_2}{k_1} \sum_{u=1}^{k_1} (k_1 - u)\rho_{u0} \Big].$$

The derivation of (18) may be compared with that of (15).

6. Effect of alignment. We can examine the effect of alignment either by an examination of the values of the variance of different samples, or by the direct use of (13). For random and stratified random sampling, the effect of alignment is to increase the variance of the sample by an amount

$$\Sigma\Sigma\, a_{uv}(\rho_{0v} - \rho_{uv}) + \Sigma\Sigma\, b_{uv}(\rho_{u0} - \rho_{uv}) \qquad \text{where } a_{uv} \geq 0,\ b_{uv} \geq 0.$$

This will be positive for monotonic decreasing correlation functions, and for the majority of functions realised in practice. Thus alignment will usually increase the variance for random and stratified random samples.

For systematic samples, the position is more complicated, but, roughly, the variance is increased by an amount

$$\Sigma\Sigma\, a_{uv}(\rho_{k_1 u, k_2 v} - \bar{\rho}_{k_1 u, k_2 v}),$$

where $a_{uv} \geq 0$ and $\bar{\rho}_{k_1 u, k_2 v}$ is a mean over a rectangle, centre $\rho_{k_1 u, k_2 v}$ for u and v non-zero, and is a mean over a line, length k_1 centre $\rho_{0,\, k_2 v}$ for u zero, (and similarly for v zero). Whether this is positive or negative will depend on the correlation function, and it will have to be investigated for the types of correlation function which are encountered.

7. Limiting forms. For a continuous process, when n_1 and n_2 are large, we may, in the same manner as for linear sampling, obtain integral approximations to the sampling variance, provided that $\Sigma\Sigma\, \rho_{d_1 u, d_2 v}$ converges.

We thus have

$$\sigma^2(r_0\, r_0) = \sigma^2(st_0\, r_0) \sim \sigma^2/n_1 n_2\,, \tag{19}$$

$$\sigma^2(r_1\, r_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 + \frac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v\right]. \tag{20}$$

$$\sigma^2(r_1\, r_1) \sim \frac{\sigma^2}{n_1 n_2}\left[1 + \frac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v + \frac{2}{d_1}\int_0^\infty \rho_{u0}\,\delta u\right], \tag{21}$$

$$\sigma^2(st_1\, r_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 - \frac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1} (d_1 - |u|)\rho_{uv}\,\delta u + \frac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v\right], \tag{22}$$

$$\sigma^2(st_0\, st_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 - \frac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1} (d_1 - |u|)(d_2 - |v|)\rho_{uv}\,\delta u\,\delta v\right], \tag{23}$$

$$\sigma^2(st_1\, st_1) \sim \frac{\sigma^2}{n_1 n_2}\Bigg[1 - \frac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{uv}\,\delta u\,\delta v$$
$$- \frac{1}{d_1 d_2^2}\int_{-d_2}^{d_2}\int_{-\infty}^{\infty}(d_2 - |v|)\rho_{uv}\,\delta u\,\delta v + \frac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1}(d_1 - |u|)$$
$$\cdot\,(d_2 - |v|)\rho_{uv}\,\delta u\,\delta v + \frac{2}{d_1}\int_0^{\infty}\rho_{u0}\,\delta u - \frac{2}{d_1^2}\int_0^{d_1}(d_1 - u)\rho_{u0}\,\delta u$$
$$+ \frac{2}{d_2}\int_0^{\infty}\rho_{0v}\,\delta v - \frac{2}{d_2^2}\int_0^{d_2}(d_2 - v)\rho_{0v}\,\delta v\Bigg], \tag{24}$$

$$\sigma^2(sy_1\, r_0) \sim \frac{\sigma^2}{n_1 n_2}\Bigg[1 - \frac{1}{d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{uv}\,\delta u\,\delta v$$
$$+ \frac{1}{d_2}\sum_{u=-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{d_1 u, v}\,\delta v - \frac{2}{d_2}\int_0^{\infty}\rho_{0v}\,\delta v\Bigg], \tag{25}$$

$$\sigma^2(sy_1\, sy_1) \sim \frac{\sigma^2}{n_1 n_2}\Bigg[\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty}\rho_{d_1 u, d_2 v} - \frac{1}{d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{uv}\,\delta u\delta v\Bigg], \tag{26}$$

$$\sigma^2(sy_0\, sy_0) \sim \frac{\sigma^2}{n_1 n_2}\Bigg[1 - \frac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{uv}\,\delta u\delta v$$
$$- \frac{1}{d_1 d_2^2}\int_{-d_2}^{d_2}\int_{-\infty}^{\infty}(d_2 - |v|)\rho_{uv}\,\delta u\delta v$$
$$+ \frac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1}(d_1 - |u|)(d_2 - |v|)\rho_{uv}\,\delta u\delta v$$
$$+ \frac{1}{d_2^2}\sum_{u=-\infty}^{\infty}\int_{-d_2}^{d_2}(d_2 - |v|)\rho_{d_1 u, v}\,\delta v - \frac{1}{d_2^2}\int_{-d_2}^{d_2}(d_2 - |v|)\rho_{0v}\,\delta v$$
$$+ \frac{1}{d_1^2}\sum_{v=-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{u, d_2 v}\,\delta u - \frac{1}{d_1^2}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{u0}\,\delta u\Bigg]. \tag{27}$$

8. Particular case where $\rho_{uv} = \rho_u\rho_v$. We note that, if $\rho_{uv} = \rho_u\rho_v$[5] most of these forms can be simplified greatly. If we write

$$sy_u = 1 - \frac{2}{d_1}\int_0^{\infty}\rho_u\,\delta u + 2\sum_{u=1}^{\infty}\rho_{d_1 u},$$
$$st_u = 1 - \frac{2}{d_1^2}\int_0^{d_1}(d_1 - u)\rho_u\,\delta u,$$

with similar forms for sy_v and st_v, and, also

$$f_1 = \frac{2}{d_2}\int_0^{\infty}\rho_v\,\delta v, \qquad f_1' = \frac{2}{d_2^2}\int_0^{d_2}(d_2 - v)\rho_v\,\delta v, \qquad f_1'' = 2\sum_{v=1}^{\infty}\rho_{d_2 v},$$
$$f_2 = \frac{2}{d_1}\int_0^{\infty}\rho_u\,\delta u, \qquad f_2' = \frac{2}{d_1^2}\int_0^{d_1}(d_1 - u)\rho_u\,\delta u, \qquad f_2'' = 2\sum_{u=1}^{\infty}\rho_{d_1 u},$$

[5] A sufficient condition for this to be a valid autocorrelation function is that both ρ_u and ρ_v should be autocorrelation functions.

then we have, for example,

$$\sigma^2(r_1 r_0) \sim \frac{\sigma^2}{n_1 n_2}(1 + f_1), \tag{28}$$

$$\sigma^2(r_1 r_1) \sim \frac{\sigma^2}{n_1 n_2}(1 + f_1 + f_2), \tag{29}$$

$$\sigma^2(st_0 st_0) \sim \frac{\sigma^2}{n_1 n_2}(st_u st_v + st_u + st_v), \tag{30}$$

$$\sigma^2(st_1 st_1) \sim \frac{\sigma^2}{n_1 n_2}(st_u st_v + f_1 st_u + f_2 st_v), \tag{31}$$

$$\sigma^2(sy_1 sy_1) \sim \frac{\sigma^2}{n_1 n_2}(sy_u sy_v + f_1 sy_u + f_2 sy_v), \tag{32}$$

$$\sigma^2(sy_0 sy_0) \sim \frac{\sigma^2}{n_1 n_2}(st_u st_v + f_1' sy_u + f_2' sy_v). \tag{33}$$

From these we get

$$\sigma^2(st_1 st_1) - \sigma^2(sy_1 sy_1) \sim \frac{\sigma^2}{n_1 n_2} [(st_u st_v - sy_u sy_v) + f_1(st_u - sy_u) + f_2(st_v - sy_v)], \tag{34}$$

$$\sigma^2(sy_1 sy_1) - \sigma^2(st_0 st_0) \sim \frac{\sigma^2}{n_1 n_2} [\{(1 - sy_u)(1 - sy_v) - (1 - st_u)(1 - st_v)\} + f_1'' sy_u + f_2'' sy_v], \tag{35}$$

$$\sigma^2(st_0 st_0) - \sigma^2(sy_0 sy_0) \sim \frac{\sigma^2}{n_1 n_2} [f_1'(st_u - sy_u) + f_2'(st_v - sy_v)]. \tag{36}$$

The forms (34), (35) and (36) enable us to compare the variances of the samples in two dimensions by using the one-dimensional results. For most practical cases, we know that the f's are positive, $st_u \geq sy_u$ and $st_v \geq sy_v$, so that

$$\sigma^2(st_1 st_1) \geq \sigma^2(sy_1 sy_1) \geq \sigma^2(st_0 st_0) \geq \sigma^2(sy_0 sy_0). \tag{37}$$

The values of $\sigma^2(st_0 st_0)/\sigma^2(r_0 r_0)$, $\sigma^2(sy_1 sy_1)/\sigma^2(r_0 r_0)$, $\sigma^2(sy_0 sy_0)/\sigma^2(r_0 r_0)$ and $\sigma^2(st_0 st_0)/\sigma^2(sy_0 sy_0)$ for $\rho_{d_1 u} = \rho_1^{|u|}$ and $\rho_{d_2 v} = \rho_2^{|v|}$ are given in table 3. It is not difficult to show that for a given number of samples, (d_1, d_2 fixed), $\sigma^2(st_0 st_0)$, $\sigma^2(sy_1 sy_1)$ and $\sigma^2(sy_0 sy_0)$ are least when $\rho_1 = \rho_2$. The expressions tabulated have a value of 1 for $\rho_1 = \rho_2 = 0$ and tend to limiting values of 0, 2/3, 0, and 2 respectively as ρ_1 and ρ_2 tend to 1. It is interesting to note that for ρ_1 and ρ_2 differing by more than 0.4 the grid imposed by $sy_1 sy_1$ is less efficient than purely random sampling. The type of function $\rho_{uv} = \rho_u \rho_v$[6] is, however, less likely to be realised

[6] For a town survey, we might find the correlation between two points depending on a within-streets and a between-streets correlation, so that this function could be realised.

TABLE 3

Comparison of the efficiencies of systematic and random sampling for various values of ρ_1 and ρ_2

ρ_2 \ ρ_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	1.000 1.000 1.000 1.000	1.000 1.222 1.000 1.000	1.000 1.500 1.000 1.000	1.000 1.857 1.000 1.000	1.000 2.333 1.000 1.000	1.000 3.000 1.000 1.000	1.000 4.000 1.000 1.000	1.000 5.667 1.000 1.000	1.000 9.000 1.000 1.000	1.000 19.000 1.000 1.000	1.000 ∞ 1.000 1.000
0.1		0.720 0.739 0.596 1.21	0.669 0.754 0.534 1.25	0.632 0.827 0.493 1.28	0.601 0.956 0.462 1.30	0.575 1.160 0.437 1.31	0.551 1.488 0.416 1.32	0.529 2.055 0.398 1.33	0.508 3.215 0.382 1.33	0.489 6.734 0.367 1.33	0.471 ∞ 0.354 1.33
0.2			0.609 0.706 0.462 1.32	0.565 0.721 0.416 1.36	0.529 0.788 0.380 1.39	0.497 0.914 0.352 1.41	0.469 1.134 0.328 1.43	0.443 1.532 0.307 1.44	0.419 2.362 0.289 1.45	0.396 4.911 0.272 1.46	0.375 ∞ 0.257 1.46
0.3				0.516 0.689 0.365 1.41	0.476 0.707 0.327 1.45	0.441 0.778 0.297 1.49	0.409 0.924 0.271 1.51	0.380 1.209 0.249 1.53	0.354 1.825 0.229 1.54	0.328 3.751 0.212 1.55	0.305 ∞ 0.196 1.55
0.4					0.432 0.680 0.288 1.50	0.394 0.702 0.256 1.54	0.360 0.787 0.229 1.57	0.329 0.983 0.206 1.60	0.300 1.437 0.185 1.62	0.272 2.900 0.167 1.63	0.247 ∞ 0.151 1.64
0.5						0.354 0.675 0.223 1.59	0.317 0.703 0.195 1.63	0.284 0.821 0.171 1.66	0.253 1.139 0.150 1.68	0.223 2.228 0.132 1.70	0.196 ∞ 0.115 1.71
0.6							0.279 0.671 0.167 1.67	0.243 0.712 0.142 1.71	0.210 0.908 0.121 1.74	0.180 1.679 0.102 1.76	0.151 ∞ 0.085 1.78
0.7								0.206 0.669 0.118 1.75	0.172 0.742 0.096 1.79	0.139 1.226 0.077 1.82	0.109 ∞ 0.059 1.84
0.8									0.136 0.667 0.074 1.84	0.102 0.863 0.055 1.87	0.070 ∞ 0.037 1.89
0.9										0.067 0.667 0.035 1.92	0.034 ∞ 0.018 1.95

in practice than a centrally-symmetric function, which is independent of the choice of axes. For this reason, we consider next this latter type of function.

9. Centrally-symmetric correlation functions. Dedebant and Wehrte [3] have given a necessary and sufficient condition for $\rho(u, v)$ to be a correlation function as

$$\rho(u, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(\omega u - \mu_v)\delta F(\omega, \mu), \tag{38}$$

or alternatively,

$$f(\omega, \mu) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(\omega u - \mu_v)\rho(u, v)\,\delta u\,\delta v. \tag{39}$$

For a centrally-symmetric correlation function we can put $u = r\cos\theta, v = r\sin\theta$ then $\rho(u, v) = \rho(r)$ and

$$f(\omega, \mu) = \frac{1}{(2\pi)^2}\int_0^{\infty}\int_0^{2\pi} \cos(r\sqrt{\omega^2 + \mu^2}\cos\theta_1)\rho(r)r\,d\theta_1\,dr,$$

where $\theta_1 = \theta + \tan^{-1}(\mu/\omega)$,

$$= \frac{1}{2\pi}\int_0^{\infty} J_0(r\tau)\rho(r)r\,dr, \quad \text{where} \quad \tau = \sqrt{\omega^2 + \mu^2}.$$

Thus, if $\rho(u, v)$ is centrally-systematic, then so is $f(\omega, \mu)$ and conversely, so that we get

$$f(\tau) = \frac{1}{2\pi}\int_0^{\infty} J_0(r\tau)\rho(r)r\delta r, \tag{40}$$

and

$$\rho(r) = 2\pi\int_0^{\infty} J_0(r\tau)f(\tau)\tau\delta\tau. \tag{41}$$

We can thus find suitable forms for $\rho(r)$ and $f(\tau)$. In this connection the formula $\int_0^{\infty} J_0(yz)e^{-ay}\delta y = 1/(a^2 + z^2)^{1/2}, a \geq 0$, is useful, since we can see that $\frac{\delta^n}{\delta a^n}(e^{-ay}/y)$ and $\frac{\delta^n}{\delta a^n}(a^2 + z^2)^{-1/2}$ are possible functions for $2\pi f(\tau)$ and $\rho(r)$ although our choice must be limited by the stochastic nature of $\rho(r)$ as well as by its convergence. Thus, for example, $a = n = 0$ gives $1/2\pi\tau$ and $1/r$ as spectral and correlation functions, but these will not converge.

In the linear case, the Markoff process $\rho(u) = e^{-au}$ had a spectral function $f(\tau) = 1/\pi(a^2 + \tau^2)$ which is a Cauchy distribution in one dimension. If we take a two-dimensional Cauchy distribution[7] as our spectral function we get $f(\tau) =$

[7] In the same way as the ordinary Cauchy distribution can be considered as a density distribution on a line produced by a point source at a distance a, radiating in all directions, so can a two-dimensional distribution be considered as a density distribution on a plane from a source at distance a.

$a/2\pi(a^2+r^2)^{3/2}$ and $\rho(r) = -\dfrac{\delta}{\delta a}(e^{-ar}/r) = e^{-ar}$. Thus it appears that a generalised Cauchy distribution will be the spectral function for a generalised Markoff process.

We can, of course, consider an "elliptical" Markoff process given by[8]

$$\rho(u, v) = \exp -\left[\frac{u^2}{a^2} - \frac{2muv}{ab} + \frac{v^2}{b^2}\right]^{1/2} \tag{42}$$

but, in what follows, to simplify the computation, m will be taken as zero, so that by changing the units in which d_1 and d_2 are measured, we will work with a process $\rho(r) = e^{-ar}$.

TABLE 4

Comparison of observed serial correlations with theoretical values obtained from a centrally-symmetric correlation function

Distance in miles	Rows		Columns		North-east		South-east	
	Observed	Calculated	Observed	Calculated	Observed	Calculated	Observed	Calculated
1	0.332	0.368	0.310	0.368	—	—	—	—
2	—	—	—	—	0.264	0.243	0.264	0.243
2	0.149	0.135	0.090	0.135	—	—	—	—
$2\sqrt{2}$	—	—	—	—	0.050	0.059	0.129	0.059
3	0.009	0.050	−0.029	0.050	—	—	—	—
$3\sqrt{2}$	—	—	—	—	−0.050	0.018	0.070	0.018
4	0.034	0.018	−0.041	0.018	—	—	—	—
$4\sqrt{2}$	—	—	—	—	−0.020	0.004	0.060	0.004

This process does not seem to be far removed from the type of correlation function experienced in agricultural field work.[9] Osborne [4] has mentioned the possible use of $\rho_u = e^{-\lambda u}$. Mahalanobis [5] has calculated correlations for a paddy field of 800 cells; his values are shown in table 4, together with values of the function e^{-r}. Bearing in mind that the standard error of each of Mahalanobis' values is approximately 0.035, the fit is seen to be quite good, although an elliptical process with axes running south-east and north-east would undoubtedly fit the observations better.

[8] In this light, $\rho(r) = e^{-ar}$ will be called the circular Markoff process, while $\rho_{u,v} = \rho_1^{|u|}\rho_2^{|v|}$ and $\rho_{u,v} = \exp\left\{-\left|\dfrac{u}{a} + \dfrac{v}{b}\right|\right\}$ will be known as degenerate Markoff processes of the first and second orders.

[9] This is further supported by the fact that using a function of this kind it is possible to obtain numerically a law in substantial agreement with Fairfield-Smith's law over a wide range of values.

10. The relative efficiencies of systematic and stratified random sampling. Ideally the correlation functions developed in the last section should be used in the expression (19)–(27), but these functions are not capable of easy integration. An alternative approach can be made if we note that

$$\frac{\sigma^2(st_0st_0) - \sigma^2(sy_0sy_0)}{\sigma^2(r_0r_0)} \sim \frac{1}{d_1}\int_{-d_1}^{d_1}\left(1 - \frac{|u|}{d_1}\right)F(u, d_2)\delta u + \frac{1}{d_2}\int_{-d_2}^{d_2}\left(1 - \frac{|v|}{d_2}\right)F(v, d_1)\delta v \tag{43}$$

where

$$F(u, d_2) = \frac{2}{d_2}\left[\int_0^{d_2}\frac{v}{d_2}\rho_{uv}\delta v + \int_{d_2}^{\infty}\rho_{uv}\delta v - d_2\sum_{v=1}^{\infty}\rho_{u,d_2v}\right],$$

$$F(v, d_1) = \frac{2}{d_1}\left[\int_0^{d_1}\frac{u}{d_1}\rho_{uv}\delta u + \int_{d_1}^{\infty}\rho_{uv}\delta u - d_1\sum_{u=1}^{\infty}\rho_{d_1u,v}\right].$$

It is seen that $F(u, d_2)$ and $F(v, d_1)$ are extensions of the expressions obtained for $(\sigma_{st}^2 - \sigma_{sy}^2)/\sigma_r^2$ in section 2. Hence, if $F(u, d_2)$ and $F(v, d_1)$ are both positive functions, systematic sampling is more accurate than stratified random sampling. A particular case of this occurs when $\rho_{uv} = \rho_1^u\rho_2^v$. However when $\rho_{uv} = \exp\{-(u^2 + v^2)^{1/2}\}$, $F(u, d_2)$ is not always positive, since, as u increases, ρ_{uv} becomes a convex function of v. This complicates the interpretation of (43) greatly since it appears that as u varies from 0 to d_1, $F(u, d_2)$ varies from $+\infty$ to an unknown value X. This value will be positive if $d_2 >> d_1$ and negative if $d_1 >> d_2$ so that if the sampling is disproportionate in the two directions systematic sampling will be more efficient than stratified random sampling. Furthermore, if $d_1 = d_2 = d$ and $d \to 0$, $F(u, d) \to \infty$ and systematic sampling again appears to be more efficient. Thus in a wide variety of cases this type of systematic sampling i.e. sy_0sy_0 gives a more accurate result than random sampling.

11. Estimation of sampling errors. An examination of formulas (7)–(18) shows that the principles used for the estimation of linear errors can be used in plane sampling. If we consider that each sample can be broken up into independent units each of which is situated in one of s strata, then for q replications we will have $qr - s$ degrees of freedom for error. For example, r_0r_0, r_0r_1, st_0r_0 and st_0r_1 will have $qn_1n_2 - 1$, $qn_2 - 1$, $qn_1n_2 - n_1$ and $qn_2 - 1$ degrees of freedom respectively, so that a single sample will contain an unbiased estimate of error, but st_0st_0, st_0st_1, st_1st_1, sy_0sy_0 and sy_1sy_1 will have $n_1n_2(q - 1)$, $n_2(q - 1)$, $q - 1$ and $q - 1$ degrees of freedom and will require replication to form a valid estimate of error. We can however use the method of splitting our sample into several parts each of which will give a fairly accurate estimate of error. We may, again, consider the possibility of using a set of systematic samples, which are evenly spaced, to estimate the sampling error, and we will see that the exclusion of the ρ's of lower order may lead to appreciable bias unless the correlation between

successive terms of the sample is small, but, as Yates has pointed out, this method will provide an upper limit for our sampling error. These methods of sampling are illustrated by the examples given below.

12. Examples. We shall consider the three methods of estimating the sampling errors of a systematic sample:

(1) using sets of systematic samples randomly placed with respect to each other, i.e. the material to be sampled is broken up into a series of sub-areas or blocks and several systematic samples are taken in each block; the error variance is calculated from the variances of the systematic samples in each block,

(2) using one set of systematic samples randomly placed, i.e. several systematic samples are taken and the area is then broken up into sub-areas or blocks; the error variance is calculated from the variances of the portions of the systematic samples in each block,

(3) using one systematic sample i.e. one systematic sample is taken which is broken into several systematic samples of wider spacing, e.g. four samples at four times the original spacing, the area is then divided into several sub-areas and the error variance is calculated from the variances of the portions of the sub-systematic samples in each block.

These three methods are increasingly accurate in their estimation of the mean, increasingly biased in their estimation of the sampling variance, and decreasingly difficult in their practical application, so that our method of sampling may vary according to the population and according to the use to which the results are to be put. It is, for example, conceivable that subsequent sampling will yield an improved estimate of error so that initially only a rough guide may be required.

a. If we are sampling from a continuous linear population with a large number of observations in each part into which we split our series, methods (1) and (2) will both give accurate estimates of the variance per term

$$\sigma^2\left(1 - \frac{2}{d}\int_0^\infty \rho_u\,\delta u + 2\sum_{u=1}^{\infty}\rho_{du}\right).$$

Method (3) will, however, estimate σ^2 instead of the correct variance per term, which is

$$\sigma^2\left(1 - \frac{2q}{d}\int_0^\infty \rho_u\,\delta u + 2\sum_{u=1}^{\infty}\rho_{du/q}\right).$$

Thus the estimates of sampling variance by method (3) will in general be higher than the estimates by methods (1) and (2), although the actual variance will be lower.

b. Kendall [6, 7] has constructed 480 terms of an artificial series $u_{n+2} = 1.1\,u_{n+1} - 0.5\,u_n + \epsilon_{n+2}$ where the ϵ_n are rectangularly distributed from -49 to 49. For this series $\sigma^2 = 2379.81$ and $s^2 = 2535.11$. The series was split in six parts of 80 terms, for each of which $n = 5$, $k = 16$, $q = 4$, so that 18 degrees of freedom were available for error. The results for this sampling configuration are

given in table 5. The values in this table corroborate the conclusions for large samples of continuous populations.

c. A number of uniformity trials were taken and sampled according to the systems st_1st_1 and sy_1sy_1. For sampling according to the system st_1st_1 the error

TABLE 5

Comparison of three methods of estimating the sampling error of systematic samples for an autoregressive scheme

Method	Estimate of sampling variance per term, s^2, based on 18 degrees of freedom	E (s^2)	True sampling variance per term
(1)	3228	2170	2170
(2)	1872	2170	2167
(3)	3709	2577	423

TABLE 6

Comparison of efficiencies of different methods of sampling on three uniformity trials

Source	Kalamkar [8]			Wiebe [9]			Wynne Sayes and Karishna, Iyez [10]		
No. In Cochran's [11] Catalogue	72			132			108		
Crop	Potatoes			Wheat			Sugar cane		
No. of Plots	576			1440			960		
Mean	23 262			587 95			270.89		
Variance per term	15 555			10,918 0*			1794 42		
Type of sampling	st_1 st_1	sy_1 sy_1	sy_1 sy_1	st_1 st_1	sy_1 sy_1	sy_1 sy_1	st_1 st_1	sy_1 sy_1	sy_1 sy_1
Proportion sampled	1/6	1/6	1/6	1/9	1/9	1/9	1/8	1/8	1/8
Method of estimating error		(2)	(3)		(2)	(3)		(2)	(3)
No. of partitions	1	4	4	1	4	4	1	5	5
n_1	3	3	6	4	2	4	4	2	4
k_1	2	2	1	3	6	3	2	4	2
n_2	16	2	4	20	5	10	15	3	6
k_2	6	12	6	6	6	3	8	8	4
q	2	4	1	2	4	1	2	4	1
Mean	23.140	23 435	23 323	586.54	598.65	275.29	275.29	266.72	271.27
Estimated variance per term	9.763	2 689	4.889	5151.6	5772.7	7038 5	1320 15	799.29	1269.54
Degrees of freedom of estimated variance	48	12	12	80	12	12	60	15	15

* Based on the original 1500 plots.

was estimated by taking two samples per strata, while, for sampling according to the system sy_1sy_1, the error was estimated by comparing sets of four samples in each part of the series by methods (2) and (3). The results of this sampling are shown in table 6. While the number of trials is small, the trend to be seen in the results agrees very well with the conclusions reached above.

13. Trend in the population. Frequently in taking samples from a population, we are faced with the problem of a trend. This will not greatly affect random and stratified random samples as estimates of the population mean, but the efficiency of systematic samples will be affected to a large extent. If we consider linear sampling, and denote by S_i the sample whose first element is x_i then the set of samples S_i will usually be monotonic with i and the difference between S_1 and S_k will be large (roughly equal to $x_1 - x_k$).

Yates [1] has suggested a method to overcome this difficulty; by letting S_i represent

$$\frac{1}{n-1}\left[\frac{i}{k}x_i + x_{i+k} + \cdots + x_{i+(n-2)k} + \frac{k-i}{k}x_{n+(n-1)k}\right],$$

the difference between systematic samples due to trend is largely removed. It is easily seen that this necessitates a small loss of information, and in particular, for a continuous random population the variance is $(n - \frac{3}{2})\sigma^2/(n-1)^2$ instead of σ^2/n. For plane samples, the corresponding adjusted sample will be

$$\begin{aligned} S_{ij} = \frac{1}{(n_1-1)(n_2-1)}\Bigg[&\frac{ij}{k_1k_2}x_{ij} + \frac{j}{k_2}x_{i+k_1,j} + \cdots + \frac{j(k_1-i)}{k_1k_2}x_{i+(n_1-1)k_1,j} \\ &+ \frac{i}{k_1}x_{i,j+k_2} + x_{i+k_1,j+k_2} + \cdots + \frac{(k_1-i)}{k_1}x_{i+(n_1-1)k_1,j+k_2} \\ &+ \frac{i(k_2-j)}{k_1k_2}x_{i,j+(n_2-1)k_2} + \cdots + \frac{(k_1-i)(k_2-j)}{k_1k_2}x_{i+(n_1-1)k_1,j+(n_2-1)k_2}\Bigg] \end{aligned}$$

with a similar loss of information.

Trend is, however, most likely to be appreciable in large samples, and in this case, the loss of information due to end adjustments is negligible, so that the conclusions reached above will remain unaltered.

The author wishes to thank Dr. F. Yates and Professor M. S. Bartlett for advice in the preparation of this paper.

REFERENCES

[1] W. G. Cochran, "The relative accuracy of systematic and stratified random samples for a certain class of populations," *Annals of Math. Stat.*, Vol. 17 (1946), p. 164.

[2] F. Yates, "A review of recent statistical developements in sampling and sampling surveys," *Roy. Stat. Soc. Jour.*, Vol. 109 (1946), p. 12.

[3] G. Dedebant and P. Wehrte, "Mecanique aliatoir," *Portugaliae Physics*, Vol. 1 (1945).

[4] J. G. Osborne, "Sampling errors of systematic and random surveys of cover-type areas," *Am. Stat. Asso. Jour.*, Vol. 37 (1942), p. 256.

[5] P. C. Mahalanobis, "On large-scale sample surveys," *Roy. Soc. Phil. Trans.*, B. 231 (1944), p. 329.

[6] M G. Kendall, "On the analysis of oscillatory time-series," *Roy. Stat. Soc Jour.*, Vol. 108 (1945), p. 93

[7] M. G. Kendall, *Contributions to the Study of Oscillatory Time-Series*, Nat. Inst. Econ. Soc. Res., 1946.

[8] R. J. KALAMKAR, "Experimental errors and the field-plot technique with potatoes," *Jour. Agr. Sci.*, 1932, p. 373.
[9] G. A. WIEBE, "Variation and correlation in grain-yield among 1500 wheat nursery plots," *Jour. Agr. Res.*, 1935.
[10] WYNNE SAYER AND P. V. KRISHNA IYER, "On some of the factors that influence the error of field experiments with special reference to sugarcane," *Ind. Jour Agr Sci.*, 1936, p. 917.
[11] W. G. COCHRAN, "Catalogue of uniformity trial data," *Roy. Stat. Soc. Suppl. Jour.*, Vol. 4 (1937), p. 233.
[12] F. YATES, "Systematic sampling," *Roy. Soc. Phil. Trans.*, Vol. 241 (1948), p. 345.

REPRESENTATION OF PROBABILITY DISTRIBUTIONS BY CHARLIER SERIES*

By R. P. Boas, Jr.

Brown University

Summary. The paper describes some results concerning the representation of a function by linear combinations of the successive differences of the Poisson distribution, not necessarily the partial sums of the type B series of Charlier.

1. Introduction. For various purposes it is often desired to expand a probability distribution $f(x)$ in a series

$$f(x) \sim \sum_{k=0}^{\infty} c_k \theta_k(x), \tag{1}$$

where the $\theta_k(x)$ are a given set of standard functions. Arguments of a heuristic nature led Charlier [4, 5, 6] to suggest that it would be useful to take the $\theta_k(x)$ in (1) to be either the successive derivatives or the successive differences of some fixed function; the two cases are often referred to as type A series and type B series, respectively. Charlier gave formulas for determining the coefficients in the two cases, but the question of whether the formal series represents the given function in any reasonable sense has to be investigated separately for each particular choice of the function generating the series. Only one special case of each type has been much used: for the A-series, $\theta_0(x)$ is the normal density function $(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$; for the B-series, $\theta_0(x)$ is the Poisson function $e^{-\lambda}\lambda^x/x!$ (when x is restricted to take only nonnegative integral values). We shall refer only to these special cases when we speak of A- and B-series in this paper.

There are two distinct problems (which have, however, often been confused) connected with the representation of a function $f(x)$ by a series (1); for convenience, we shall refer to them in this paper as the practical problem and the theoretical problem. In the *practical problem*, we have an empirical function $f(x)$, defined only for a finite number of values of x, which we suspect is representable by $c_0\theta_0(x)$ together with a small correction, so that we hope that a few (say three or four) terms of (1) may give a good representation of $f(x)$ in a relatively simple analytical form with a reasonable amount of computational labor. In some cases, and certainly with the classical A- and B-series which we are considering, we could represent, as closely as desired, any $f(x)$ (however irregular) which takes nonzero values at only a finite number of points; but there is no interest in doing this if the process involves finding too many terms of the series. (Neglect of this fact has led to ill-founded statements by mathematicians about the satisfactory nature of the A- or B-series; but see [27, pp. 38–39].)

Thus it would be of interest to know, if possible, under what circumstances a given empirical density can be represented fairly well by a few terms of a series of a given kind. If no simple criterion can be given, it is desirable to have a means

* Address delivered by invitation at the meeting of the Institute at Boulder, Colorado, on September 1, 1949.

of computing coefficients which will make a few terms of (1) give the best possible fit—best possible being defined in a way appropriate for the problem at hand.

In the *theoretical problem*, $f(x)$ is a function defined for all values of x, or at least for all of an infinite set of equally spaced values of x, arising from theoretical considerations which suggest $c_0\theta_0(x)$ as a reasonable first approximation to $f(x)$. For example, the central limit theorem states that under certain conditions the cumulative distribution function of the sum of a large number of independent random variables is approximately normal; then we might expect that this distribution function would be representable by a series (1) with $\theta_0(x)$ the normal distribution function. For such theoretical purposes we should like to have criteria for the representability of a sufficiently general $f(x)$ by a series (1), where representability is of course to be interpreted appropriately, as ordinary convergence, uniform convergence, convergence in mean square, asymptotic representation, etc., according to the requirements of the problem at hand. The larger the class of $f(x)$ for which we can prove a representation theorem, the larger is the possible domain of applicability of the series to theoretical problems.

2. The A-series. This paper is concerned with the B-series, but for comparison we first mention some properties of the A-series. In the case of the classical A-series, we have the attractive fact that the functions $\theta_n(x)$ are orthogonal with weight function $e^{\frac{1}{2}x^2}$, that is,

$$\int_{-\infty}^{\infty} \theta_n(x)\theta_m(x)e^{\frac{1}{2}x^2}\,dx = 0, \qquad m \neq n.$$

In fact, $e^{\frac{1}{2}x^2}\theta_n(x)$ is, except for a numerical factor, the nth Hermite polynomial. This orthogonality property enables one to compute the coefficients in a series (1) with great ease from

$$n!\,c_n = \int_{-\infty}^{\infty} f(x)\theta_n(x)e^{\frac{1}{2}x^2}\,dx, \tag{2}$$

or since $\theta_n(x)e^{\frac{1}{2}x^2}$ is a polynomial, from the moments of $f(x)$. By the classical theory of orthogonal functions, this means that if the c_n are so computed, and we take $N + 1$ terms of the series, we minimize

$$\int_{-\infty}^{\infty} e^{\frac{1}{2}x^2}[f(x) - F_N(x)]^2\,dx \tag{3}$$

for all possible sums

$$F_N(x) = \sum_{n=0}^{N} c_n\theta_n(x). \tag{4}$$

The convergence theory of Hermite series has been thoroughly investigated by mathematicians, so that it would appear that in theoretical problems, in which $f(x)$ is given for all values of x, we are in a position to find out everything about the representation of $f(x)$ by an A-series. Also in problems of practical curve-

fitting, the fact that the closest approximation to $f(x)$ (in the sense (3)) by sums of the form (4) is given by choosing the coefficients according to (2) seems to leave no more to be said.

However, the formal elegance of the A-series seems to be somewhat misleading. Even when a series converges it by no means follows that its Nth partial sum is the best selection of N terms for representing a given function. Even though the partial sums do give the best fit in the sense of (3), it may not be desirable to measure the closeness of approximation by (3); some other measure of approximation may be better suited to the end in view. For example, it is known that the partial sums of Edgeworth's series (see [8]), which is a rearrangement of the A-series, are more satisfactory for some purposes than the partial sums of the A-series with the coefficients determined by (2). More precisely, Edgeworth's series furnishes an asymptotic expansion, with a remainder term whose order of magnitude can be estimated quite precisely, in circumstances where the series of orthogonal functions does not do this. Again, for practical purposes a few terms of the A-series sometimes exhibit undesirable properties (such as negative frequencies). If $f(x)$ is a function defined only for integral values of x, A. Fisher [10] has suggested and applied the idea of minimizing, not (3), but the sum $\sum_{-\infty}^{\infty} |f(x) - F_n(x)|^2$ in order to determine the coefficients of the approximating sums.

3. The B-series. We can now see how the status of the B-series resembles or differs from that of the A-series. Here we deal principally with a function defined for integral values of x; $\theta_0(x) = \theta(x) = e^{-\lambda}\lambda^x/x!$, $\Delta\theta(x) = \theta(x) - \theta(x-1)$, $\Delta^k\theta(x) = \Delta(\Delta^{k-1}\theta(x))$ and $\theta_k(x) = \Delta^k\theta(x)$; $\theta(x)$ is taken to be 0 for negative integral x. We shall refer to this as the *discrete case* of the B-series. The literature of the subject contains a number of rather painful attempts to put the coefficients into usable form, persisting even after the simple formula

$$c_n = (1/n!) \sum_{l=0}^{n} \binom{n}{l} (-1)^l \lambda^{n-l} \mu_l \tag{5}$$

had been obtained, where μ_n is the nth factorial moment,

$$\mu_n = \sum_{k=n}^{\infty} f(k)k!/(k-n)!.$$

Formula (5) can be derived, for example, by using orthogonality properties of the $\theta_r(x)$. We have, in fact, that $\sum_{x=0}^{\infty} \theta_n(x)\theta_m(x)/\theta_0(x)$ is 0 or $n!\,\lambda^{-n}$ according as $n \neq m$ or $n = m$.

The parameter λ in the B-series is at our disposal, and can for example be chosen in such a way as to improve the convergence of the series. For purposes of practical curve-fitting, it has been customary to choose λ equal to the mean of the distribution $f(x)$, a choice which makes the coefficient c_1 of $\Delta\theta$ equal to zero. Charlier also suggested other methods in which c_1 and c_2, or c_1, c_2 and c_3 are zero [7]. Such choices, of course, may reduce the amount of computation needed

to make use of a given number of differences in fitting a curve; aside from this consideration their use seems to depend on the belief that one improves the convergence of a series by adjusting any available parameters so that as many as possible of the initial terms of the series are zero. This belief does not always seem to be confirmed by the facts. (In particular, compare columns 2 and 5 of Table 1, columns 2 and 4 of Table 2, or columns 2 and 4 of Table 3.)

The theoretical problem of what $f(x)$ can be represented by convergent B-series has been studied by several authors [12, 13, 17, 19, 20, 21, 23, 24, 26, 28]; the study by Schmidt [24; see also 25 and 17] gives necessary and sufficient conditions for the representation in the case of a nonnegative $f(x)$, so that, at least in all cases of interest in statistics, the theoretical problem seems to be completely solved. However, one of the purposes of the present paper is to reopen this apparently closed problem.

There is also a *continuous* version of the B-series, which is suggested by the fact that

$$\theta(x) = (2\pi)^{-1} e^{-\lambda} \int_{-\pi}^{\pi} e^{-ixu} \exp(\lambda e^{iu})\, du \tag{6}$$

reduces to the Poisson function $e^{-\lambda}\lambda^x/x!$ for positive integral x (and to 0 for negative integral x). This form of the B-series has not been much used, and its use is subject to suspicion since it has rather peculiar properties. In particular, it cannot represent, in any reasonable sense, a positive function $f(x)$ or one which is too small as $x \to \infty$ [26, 3]; since the functions which present themselves for representation in practice are both positive and small at infinity, the continuous case of the B-series looks unpromising for applications. (See also [27a], 1a.) However, it has been applied [15].

The purpose of this paper is to describe some results on the B-series which have been obtained in a mathematical paper [3], devoted to what we have called the theoretical problem; some contributions to the practical problem will also be given in the present paper. The starting point of this investigation was the question of what happens if one tries to approximate a function, not by the partial sums of the series (1), but by some other combination of the first N functions $\theta_n(x)$, when approximation is taken in the sense of (unweighted) least-squares. This method of approximation seems well adapted to statistical problems, and leads to simpler mathematical work than ordinary point-by-point convergence of the partial sums. The B-series itself gives a least squares approximation with a weight function $1/\theta_0(x)$. We consider here only the classical B-series, when $\theta_0(x) = \theta(x) = e^{-\lambda}\lambda^x/x!$, $\theta_n(x) = \Delta^n\theta_0(x)$; the main results are substantially the same for rather more general cases [3; see also 14, 25] In addition, here we consider only nonnegative $f(x)$, assumed zero for negative x. Functions which need not be zero for negative x are handled easily by generalizing the B-series to the form [3]

$$f(x) \sim \sum_{-\infty}^{-1} b_n \nabla^n \theta(x) + \sum_{0}^{\infty} a_n \Delta^n \theta(x), \tag{7}$$

where ∇ denotes the advancing difference: $\nabla\theta(x) = \theta(x) - \theta(x+1)$; there seems to be no particular reason (other than a historical one) for preferring one kind of difference to the other. The generalized series (7) might be useful for graduating symmetrical probability distributions, although it does not seem to have been considered in the literature (cf. [1a]).

4. Results: practical problem. Our question takes somewhat different forms in the two cases which we have described as the practical and the theoretical. In the former, we ask what the coefficients $a_n^{(N)}$ should be so that

$$\sum_{x=0}^{\infty} \left| f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) \right|^2 \tag{8}$$

shall be a minimum, where $f(x)$ is an empirically given function and N is a given integer, in general not very large. If N is 0, 1 or 2, that is, if we use 1, 2 or 3 terms, the best choice of the $a_k^{(N)}$ in (8) can be calculated without difficulty.

For $N = 0$, our question is that of finding the best least-squares fit to $f(x)$ by a Poisson distribution $a_0^{(0)} e^{-\lambda}\lambda^x/x!$; the best choice of $a_0^{(0)}$ is then

$$a_0^{(0)} = \left\{ e^{\lambda} \sum_{x=0}^{\infty} f(x)\lambda^x/x! \right\} \Big/ J_0(2i\lambda), \tag{9}$$

where

$$J_0(iy) = 1 + y^2/(2!)^2 + y^4/(4!)^2 + \cdots$$

(J_0 denotes the Bessel function of order 0); on the other hand, the usual formula (5) gives the different coefficient

$$c_0 = \mu_0 = \sum_{x=0}^{\infty} f(x).$$

This, of course, is simpler than (9) to compute, although its use is based on the uncritical assumption that the first term of the series (1) is the best one to take if only one term is to be used. Charlier [7; see also 10, pp. 101–103] suggested a different formula in which one uses, not $\Delta^k\theta(x)$, but $\Delta^k\theta(px+q)$, the parameters p, q, λ being adjusted to make the terms of (1) in $\Delta\theta$, $\Delta^2\theta$, $\Delta^3\theta$ all zero; here $\theta(x)$ is defined when x is not an integer by interpreting $e^{-\lambda}\lambda^x/x!$ as $e^{-\lambda}\lambda^x/\Gamma(x+1)$, and *not* by using formula (6). Table 2 shows that in at least one numerical case (9) gives a better least-squares fit than Charlier's method (and without introducing gamma functions to take care of $\theta(x)$ for fractional x). However, it is not excluded that Charlier's method will give better results in other cases, since with the change of the functions $\theta_n(x)$ the results of this paper cease to apply.

For $N = 1$, we get the best least-squares approximation to $f(x)$ by

$$a_0^{(1)}\theta(x) + a_1^{(1)}\Delta\theta(x)$$

if

$$
\begin{aligned}
a_0^{(1)} &= \frac{e^{2\lambda}}{\alpha+\beta}\,(\Sigma_0+\Sigma_1),\\
a_1^{(1)} &= \left\{\frac{\beta}{\alpha^2-\beta^2}\,\Sigma_0-\frac{\alpha}{\alpha^2-\beta^2}\,\Sigma_1\right\}e^{2\lambda},
\end{aligned}
\tag{10}
$$

where $\Sigma_0 = \sum_{x=0}^{\infty} f(x)\theta(x)$, $\Sigma_1 = \sum_{x=0}^{\infty} f(x)\theta(x-1)$, $\alpha = J_0(2i\lambda)$, $\beta = -iJ_1(2i\lambda)$, the J's again denoting Bessel functions. For $N = 2$, the corresponding formulas involve also $\gamma = -J_2(2i\lambda)$ and $\Sigma_2 = \sum_{x=0}^{\infty} f(x)\theta(x-2)$. They are:

$$
\begin{aligned}
e^{-2\lambda}a_0^{(2)} &= \frac{\beta-\alpha}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_0+\frac{2\beta-\alpha-\gamma}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_1+\frac{\beta-\alpha}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_2,\\
e^{-2\lambda}a_1^{(2)} &= \frac{\beta\gamma-\alpha\beta+2\beta^2-2\alpha\gamma}{(\alpha-\gamma)(2\beta^2-\alpha^2-\alpha\gamma)}\,\Sigma_0+\frac{\alpha+\gamma-2\beta}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_1\\
&\qquad+\frac{2\alpha^2-2\beta^2+\beta\gamma-\alpha\beta}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_2,\\
e^{-2\lambda}a_2^{(2)} &= \frac{\alpha\gamma-\beta^2}{(\alpha-\gamma)(2\beta^2-\alpha^2-\alpha\gamma)}\,\Sigma_0+\frac{\beta}{2\beta^2-\alpha^2-\alpha\gamma}\,\Sigma_1\\
&\qquad+\frac{\beta^2-\alpha^2}{(\alpha-\gamma)(2\beta^2-\alpha^2-\alpha\gamma)}\,\Sigma_2.
\end{aligned}
\tag{11}
$$

The functions $i^n J_n(iy)$ are real for real y, and extensive tables are available [32].

Some numerical examples showing the comparison between graduation by these formulas and by the corresponding number of terms of the B-series are given in Tables 1–3. It will be noticed that (as the theory indicates) one gets a better least-squares fit by formulas (9), (10) or (11) than by a corresponding number of terms of the B-series using the coefficients (5). However, one may not get a better fit if goodness of fit is measured in some other way, e.g. by χ^2. Unfortunately the coefficients calculated by this method increase rapidly in complexity as the number of terms increases, and even the coefficients for $N = 3$ would involve very heavy algebra. Since numerical examples [2] indicate that it is often necessary to go to terms in $\Delta^4\theta$ for a satisfactory fit, it might be worth while to calculate the next few coefficients.

5. Results: theoretical problem. In the case of a theoretical distribution we ask how coefficients should be determined so that

$$
\sum_{x=0}^{\infty}\left|f(x)-\sum_{k=0}^{N}a_k^{(N)}\,\Delta^k\theta(x)\right|^2 \tag{12}
$$

will tend to 0 as $N \to \infty$. The convergence to 0 of (12) is a rather strong kind of convergence, since it implies convergence of the approximating sums to $f(x)$, not only for each x, but even uniformly for all x. Of course, the "best" choice of

$a_k^{(N)}$ as above would be expected to give convergence under the weakest hypotheses, but because of the complexity of these coefficients it seems desirable to make (12) only approximately a minimum; this actually makes no difference in the limit, although the approximation is not usually satisfactory for small values of N. To see the connection between the formulas used here and the "classical" formula (5) for the coefficients in (1), we note that (5) can be written

$$a_n = \frac{(-1)^n}{n!}\sum_{k=0}^{\infty} f(k)\,\frac{d^n}{dz^n}\,[z^k\,e^{\lambda(1-z)}]_{z=1}\,; \tag{13}$$

(5) results if we expand the derivative by Leibniz's rule and rearrange the sum. If we expand $e^{-\lambda z}$ in a power series before differentiating in (13), we obtain

$$a_n = (-1)^n\sum_{k=0}^{\infty} f(k)\sum_{l=\max(k,n)}^{\infty}\binom{l}{n}e^{\lambda}(-\lambda)^{l-k}/l! = e^{\lambda}(-1)^n\sum_{l=n}^{\infty}\binom{l}{n}\sum_{k=0}^{l}\frac{(-\lambda)^{l-k}}{(l-k)!}f(k).$$

If now we break this series off at $n = N$ to obtain

$$a_n^{(N)} = e^{\lambda}(-1)^n\sum_{l=n}^{N}\binom{l}{n}\sum_{k=0}^{l}\frac{(-\lambda)^{l-k}}{(l-k)!}\,f(k), \tag{14}$$

we obtain a sequence of approximations to $f(x)$ by sums $\sum_{k=0}^{N} a_k^{(N)}\Delta^k\theta(x)$ which has, in general, much better convergence properties than the partial sums of the B-series with coefficients a_n given by (5). In particular, if $f(x) = 0$ for $x = -1, -2, \cdots$, this sequence of approximations converges to $f(x)$ whenever $\sum_{x=0}^{\infty}|f(x)|^2$ converges; on the other hand, for nonnegative $f(x)$ it is known [24] that the B-series converges if and only if $\lim_{x\to\infty} f(x)2^x x^k = 0$ for $k = 0, 1, 2, \cdots$, a much more restrictive condition. If we demand that the partial sums of the B-series converge in mean square, that is, that (12) tends to zero with $a_k^{(N)}$ independent of N, we have the even more restrictive condition [3] that $\limsup_{x\to\infty}\{f(x)\}^{1/x} \leqq \frac{1}{3}$.

The approximating sums with coefficients (14) have the additional property that they reproduce $f(x)$ exactly for $x = 0, 1, 2, \cdots, N$. One would expect that in general they would then tend to deviate rather widely from $f(x)$ for larger x, and so would not be satisfactory for practical curve-fitting. However, it seems possible that if we fit such a sum not to $f(x)$, but to $f(px + q)$, with suitable integers p and q, thus making the approximation agree with $f(x)$ at a set of values covering the whole range of definition of $f(x)$, it might give a satisfactory fit elsewhere. This possibility has not been investigated; a similar approach using the partial sums of the B-series was suggested by Charlier [7] and Fisher [10].

6. The continuous case of the B-series. In the continuous case we again ask, not when

$$f(x) = \sum_{n=0}^{\infty} a_n\Delta^n\theta(x) \tag{15}$$

with uniform convergence in every finite interval, but when

$$f(x) = \text{l.i.m.}_{N\to\infty} \sum_{n=0}^{N} a_n^{(N)} \Delta^n \theta(x), \tag{16}$$

which means that

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} \left| f(x) - \sum_{n=0}^{N} a_n^{(N)} \Delta^n \theta(x) \right|^2 dx = 0. \tag{17}$$

For (15) the following negative results are known [26]: if $f(x) \geqq 0$, (15) cannot converge uniformly on every finite interval (unless $f(x) \equiv 0$); the series, if convergent uniformly on every finite interval, cannot converge to $f(x)$ unless the Fourier transform of $f(x)$ vanishes outside $(-\pi, \pi)$, a condition which

TABLE 1

Number of petals on buttercups. $\lambda = .631$

x	1 Observed frequency	2 Calculated 3 terms (formula 5)	3 Calculated 1 term (formula 9)	4 Calculated 2 terms (formula 10)	5 Calculated 3 terms (formula 11)	6 Calculated 3 terms (formula 14)
5	133	134.9	119.9	130.6	132.9	133.0
6	55	51.6	75.6	62.3	55.3	55.0
7	23	22.5	22.5	13.3	22.1	23.0
8	7	9.5	5.0	1.5	8.5	9.1
9	2	2.9	0.8	0.0	2.4	2.6
10	2	0.6	0.1	0.0	0.5	0.5
Total.	222	222.0	223.9	207.7	221.7	223.2

automatically excludes any $f(x)$ which vanishes for all large $|x|$ or even is too small as $x \to \infty$. Nevertheless, Jørgensen [15] applies the continuous case successfully to practical problems. A possible explanation of this apparent discrepancy is that if the $a_n^{(N)}$ in (16) are properly determined, (16) will be true under fairly general conditions. To be sure, the mean square difference in (17) cannot be made arbitrarily small unless the Fourier transform $g(x)$ of $f(x)$ vanishes outside $(-\pi, \pi)$, but if $|f(x)|^2$ is integrable the difference can be made small if $g(x)$ is itself small. If $g(x)$ does vanish outside $(-\pi, \pi)$, then (16) is true; and in fact the coefficients $a_k^{(N)}$ can be taken the same as in (14), so that the approximating sums depend only on the values of $f(x)$ for integral values of x; these values are known to determine $f(x)$ under our hypotheses on $g(x)$.

7. Discussion of some numerical results. Table 1. Column 2 gives the fit by two terms of the B-series (really three, since the coefficient of $\Delta\theta$ is zero when

formula (5) is used), as calculated by Charlier [7] (that is, using terms through $\Delta^2\theta$). Column 3 gives the best least-squares fit by a single term, i.e., a Poisson distribution, calculated by formula (9); it is clear that this term alone does not represent the observations very well. Column 4 gives the best least-squares fit by terms through $\Delta\theta$. Column 5 gives the best least-squares fit by terms through $\Delta^2\theta$; the improvement over Charlier's fit by the same number of terms is evident by inspection. Column 6 gives, for comparison, the same number of terms calculated by formula (14), which gives an approximation to the best least-squares fit and necessarily reproduces the data exactly for the first three

TABLE 2

Failure of grains of barley. $\lambda = 2.757$

x	1 Observed frequency	2 Calculated 4 terms (Charlier)	3 Calculated 1 term (Formula 9)	4 Calculated 2 terms (Formula 10)	5 Calculated 3 terms (Formula 11)
0	53	63	47.3	49.9	48.4
1	131	139	130.4	134.7	133.4
2	180	174	179.8	181.6	182.3
3	170	151	165.3	163.2	164.3
4	111	111	113.9	110.0	109.8
5	50	60	62.7	59.3	58.1
6	22	32	28.8	26.5	25.2
7	22	14	11.4	10.2	9.3
8	7	6	3.9	3.4	2.9
9	2	2	1.1	1.0	0.8
10	1	0	0.3	0.2	0.2
Total.....	749	752	744.9	740.0	734.7

values of x. The fact that (14) gives good results here is presumably connected with the small size of λ.

Table 2. Column 2 gives the values calculated by Charlier [7] for a fit after the linear transformation $x \to px + q$, with λ, p and q chosen to make the terms in $\Delta\theta$, $\Delta^2\theta$, $\Delta^3\theta$ all zero (the values were read to the nearest integer from Charlier's graph). Column 3 gives the best least-squares single-term fit calculated by formula (9); this is a considerable improvement for $x \leqq 6$, but for the remainder of the table it is rather poor. Column 4 gives the best least-squares fit by two terms; column 5, that by three. The χ^2-test indicates that the graduation is rather poor in all cases.

Table 3. Column 2 gives the classical calculation with terms through $\Delta^2\theta$; this was given by A. Fisher [10] and (more accurately) by Aroian [2]. Columns 3

and 4 give the best least-squares approximations by two and three terms; column 4 is better than column 2, in this sense, as expected. However, column 4 is a poorer fit when tested by χ^2, chiefly because of the poor fit at $x = 0$. It should be noted that two more terms of the B-series give a more satisfactory fit [2].

TABLE 3

α-particles from a bar of polonium. $\lambda = 3.87155$

x	1 Observed frequency	2 Calculated 3 terms (formula 5)	3 Calculated 2 terms (formula 10)	4 Calculated 3 terms (formula 11)
0	57	49.5	51.3	45.2
1	203	201.3	213.3	190.9
2	383	403.4	399.0	393.5
3	525	532.3	524.8	529.8
4	532	520.6	517.2	525.4
5	408	402.6	407.7	409.7
6	273	254.8	267.7	261.9
7	139	137.1	150.6	141.1
8	45	64.0	74.1	65.3
9	27	26.1	32.4	26.3
10	10	9.4	12.8	9.3
11	4	3.0	4.6	2.9
12	0	0.9	1.5	0.8
13	1	0.2	0.5	0.2
14	1	0.0	0.1	0.0
Total.....	2608	2605.2	2657.6	2602.3
		$\chi^2 = 10.2$ $n = 7$	$\chi^2 = 16.2$ $n = 8$	$\chi^2 = 11.4$ $n = 7$

8. Proofs: theoretical problem. We now outline the proofs of the results which we have stated. They depend on the fact that the numbers $\theta(x)$ ($x = 0, \pm 1, \pm 2, \cdots$) (where $\theta(x) = 0$ when x is a negative integer) are the Fourier coefficients of the function $\varphi(u) = e^{-\lambda} \exp(\lambda e^{iu})$, i.e.

$$\theta(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} \varphi(u) e^{-ixu}\, du, \qquad x = 0, \pm 1, \pm 2, \cdots$$

Furthermore,

$$\Delta^k \theta(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} \varphi(u)(1 - e^{iu})^k e^{-ixu}\, du.$$

If we then assume the condition $\sum_{-\infty}^{\infty} |f(x)|^2 < \infty$, with $f(x) = 0$ for $x = -1, -2, \cdots$, the numbers $f(x)$ are the Fourier coefficients of a function $g(x)$ of integrable square, by the Riesz-Fischer theorem from the theory of Fourier series [31, p. 74]:

$$f(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} g(u) e^{-ixu}\, du, \qquad x = 0, \pm 1, \pm 2, \cdots.$$

Thus

$$(18)\quad f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ixu} \left[g(u) - \varphi(u) \sum_{k=0}^{N} a_k^{(N)} (1 - e^{iu})^k \right] du,$$

and so the expressions on the left appear as the Fourier coefficients of the expressions in square brackets on the right. By Parseval's theorem for Fourier series [31, p. 76], then, we have

$$(19)\quad \sum_{x=-\infty}^{\infty} \left| f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) \right|^2 = (2\pi)^{-1} \int_{-\pi}^{\pi} \left| g(u) - \varphi(u) \sum_{k=0}^{N} a_k^{(N)} (1 - e^{iu})^k \right|^2 du.$$

Thus we have reduced the problem of minimizing the mean-square difference on the left of (19) to that of minimizing the integral on the right of (19). By rearranging the sum in the integrand, we see that an equivalent problem is to minimize

$$(20)\qquad D = (2\pi)^{-1} \int_{-\pi}^{\pi} \left| g(u) - \varphi(u) \sum_{k=0}^{N} c_k^{(N)} e^{iku} \right|^2 du,$$

where the $c_k^{(N)}$ and $a_k^{(N)}$ are readily expressed in terms of each other; in fact,

$$(21)\qquad a_k^{(N)} = (-1)^k \sum_{l=k}^{N} \binom{l}{k} c_l^{(N)}.$$

Since $|\varphi(u)| = e^{-\lambda + \lambda \cos u} \geqq e^{-2\lambda} > 0$, we can write D in the form

$$D = (2\pi)^{-1} \int_{-\pi}^{\pi} \left| g(u)/\varphi(u) - \sum_{k=0}^{N} c_k^{(N)} e^{iku} \right|^2 |\varphi(u)|^2\, du,$$

so that

$$(22)\qquad \int_{-\pi}^{\pi} \left| g(u)/\varphi(u) - \sum_{k=0}^{N} c_k^{(N)} e^{iku} \right|^2 du \geqq 2\pi D \geqq e^{-4\lambda} \int_{-\pi}^{\pi} \left| g(u)/\varphi(u) - \sum_{k=0}^{N} c_k^{(N)} e^{iku} \right|^2 du,$$

since $e^{-2\lambda} \leqq |\varphi(u)| \leqq 1$. Thus we can make D arbitrarily small if and only if we can make

$$D^* = (2\pi)^{-1} \int_{-\pi}^{\pi} \left| g(u)/\varphi(u) - \sum_{k=0}^{N} c_k^{(N)} e^{iku} \right|^2 du \tag{23}$$

arbitrarily small. Now the Fourier coefficients of $g(u)$ are $f(x)$; those of $1/\varphi(u)$ are $e^{\lambda}(-\lambda)^x/x!$ for $x \geqq 0$, 0 for $x < 0$; by the convolution theorem for Fourier coefficients [31, p. 90] the nth Fourier coefficient of $g(u)/\varphi(u)$ is

$$\sum_{k=0}^{n} f(n-k) e^{\lambda}(-\lambda)^k/k!, \qquad n = 0, 1, 2, \cdots, \tag{24}$$

and zero for $n < 0$. Furthermore, it is well known from the theory of Fourier series that D^* is a minimum if $c_k^{(N)}$ are chosen as the first $N+1$ Fourier coefficients of $g(u)/\varphi(u)$, and that this minimum is arbitrarily small for large enough N if and only if the Fourier coefficients of $g(u)/\varphi(u)$ are zero for negative indices —which is in fact the case. If we then take the values (24) for $c_k^{(N)}$, $k = 0, 1, \cdots, N$, and express $a_k^{(N)}$ in terms of $c_k^{(N)}$ by (21), we arrive at the formula (14).

It will be observed that the minimum D is connected with the minimum D^* by

$$\min D \leqq \max |\varphi(u)| \cdot \min D^* \leqq \min D^* \leqq \frac{\min D}{\min |\varphi(u)|} \leqq e^{2\lambda} \min D,$$

so that all that we can say about the approximation given by (14) with a small N is that it is an upper bound for the best possible mean-square approximation by sums (18), and that the best mean-square approximation is at worst $e^{-2\lambda}$ times it This means that if D^* is small, so is D; but D^* is not necessarily small even if D is. Hence we cannot in general expect the coefficients (14) to be suitable for practical curve-fitting, since they may increase the mean-square error by a factor of as much as $e^{2\lambda}$, we may, however, expect (14) to be better when λ is small.

Now, as we have already observed,

$$f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x)$$

is the xth Fourier coefficient of

$$g(u) - \varphi(u) \sum_{k=0}^{N} a_k^{(N)} (1 - e^{it})^k ;$$

if we write (18) in the form

$$f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) = \int_{-\pi}^{\pi} e^{-ixt} \left[g(t)/\varphi(t) - \sum_{k=0}^{N} a_k^{(N)} (1 - e^{it})^k \right] \varphi(t)\, dt, \tag{25}$$

and choose the $a_k^{(N)}$ as specified above, the expression in square brackets is $g(t)/\varphi(t)$ minus the first $N+1$ terms of its Fourier series, and so the Fourier

series of $[\cdots]$ involves no e^{ikt} with $k < N + 1$. Since the Fourier series of $\varphi(t)$ involves no e^{ikt} with $k < 0$, the product $\varphi(t)[\cdots]$ also involves no e^{ikt} with $k < N + 1$, and therefore the integral in (25) is zero for $x = 0, 1, 2, \cdots, N$ (since it represents the xth Fourier coefficient of $\varphi(t)[\cdots]$). In other words,

$$f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) = 0, \qquad x = 0, 1, 2, \cdots, N.$$

Furthermore, we can compute $f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x)$ for $x > N$ by the convolution formula from the Fourier series of $\varphi(t)$ and $[\cdots]$; for $n > N$, the nth Fourier coefficient of $[\cdots]$ is just that of $g(t)/\varphi(t)$, given by (24), and that of $\varphi(t)$ is $e^{-\lambda}\lambda^n/n!$, so for $x > N$

$$f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) = \sum_{l=N+1}^{x} \left(\sum_{k=0}^{l} f(l-k) e^{\lambda} (-\lambda)^k / k! \right) \theta(x-l)$$

and in particular

$$f(N+1) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(N+1) = \sum_{k=0}^{N+1} f(N+1-k)(-\lambda)^k / k!.$$

9. Proofs: practical problem. We have so far obtained only an estimate for the minimum of D, by obtaining the minimum of D^*; this estimate is satisfactory for large N and so for theoretical purposes. However, to obtain precisely the best mean-square approximation to $f(x)$ by a small number N of terms of the sum in (18), we have to choose $a_k^{(N)}$ so that

$$\sum_{k=0}^{N} a_k^{(N)} (1 - e^{it})^k \varphi(t)$$

is the first $N + 1$ terms of the expansion of $g(t)$ in terms of the set of functions obtained by replacing $(1 - e^{it})^k \varphi(t)$, $k = 0, 1, 2, \cdots$, by an equivalent orthonormal set. The process for obtaining this orthonormal set is well known, it turns out that the integrals that have to be evaluated are expressible in terms of Bessel functions of imaginary argument; the result is that the first orthonormal functions are

$$\psi_0(t) = (2\pi)^{-\frac{1}{2}} \alpha^{-\frac{1}{2}} \exp(\lambda e^{it}),$$

$$\psi_1(t) = (2\pi)^{-\frac{1}{2}} \frac{\alpha_1 - \alpha_0 e^{it}}{[\alpha_0(\alpha_0^2 - \alpha_1^2)]^{\frac{1}{2}}} \exp(\lambda e^{it}),$$

$$\psi_2(t) = (2\pi)^{-\frac{1}{2}} \frac{\alpha_1^2 - \alpha_0\alpha_2 - \alpha_1(\alpha_0 - \alpha_2)e^{it} - (\alpha_1^2 - \alpha_0^2)e^{2it}}{[(\alpha_1^2 - \alpha_0^2)(\alpha_0 - \alpha_2)(2\alpha_1^2 - \alpha_0^2 - \alpha_0\alpha_2)]^{\frac{1}{2}}} \exp(\lambda e^{it}),$$

where $\alpha_0 = J_0(2i\lambda)$, $\alpha_1 = -iJ_1(2i\lambda)$, $\alpha_2 = -J_2(2i\lambda)$. It is then a simple matter, first to express ψ_0, ψ_1, ψ_2 in terms of $\varphi(t)$, $\varphi(t)(1 - e^{it})$, $\varphi(t)(1 - e^{it})^2$, and then to determine $a_0^{(0)}$; $a_0^{(1)}$, $a_1^{(1)}$; and $a_0^{(2)}$, $a_1^{(2)}$, $a_2^{(2)}$. For example, the best two-term

approximation for $g(u)$ in terms of $\psi_0(u)$, $\psi_1(u)$ is

$$g(u) \backsim \psi_0(u) \int_{-\pi}^{\pi} g(u)\bar{\psi}_0(u)\, du + \psi_1(u) \int_{-\pi}^{\pi} g(u)\bar{\psi}_1(u)\, du,$$

and the integrals $\int_{-\pi}^{\pi} g(u)\bar{\psi}_k(u)\, du$ are combinations of terms of the form

$$(2\pi)^{-1} \int_{-\pi}^{\pi} g(u)e^{iku}\, \varphi(u)\, du;$$

these in turn are Fourier coefficients of $g(u)\varphi(u)$ and so are expressible, by the Parseval formula, as products of the Fourier coefficients of $g(u)$ (namely, $f(n)$) and of $\varphi(u)$ (namely, $\theta(n)$) We omit the algebraic work; the results are given in formulas (9), (10), (11).

10. Proofs: continuous case. In the continuous case of our approximation problem we assume that $|f(x)|^2$ is integrable on $(-\infty, \infty)$ and look for coefficients $a_k^{(N)}$ that will minimize

$$D = \int_{-\pi}^{\pi} \left| f(x) - \sum_{k=0}^{N} a_k^{(N)} \Delta^k \theta(x) \right|^2 dx,$$

where

$$\theta(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} \varphi(u)e^{-ixu}\, du,$$

$$\Delta^k \theta(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} \varphi(u)e^{-ixu}(1 - e^{iu})^k\, du.$$

Let $f(x)$ be the Fourier transform of $g(u)$; we can regard $\theta(x)$ as the Fourier transform of $\varphi(u)$, $\varphi(u)$ being defined as zero outside $(-\pi, \pi)$. Then by Parseval's theorem for Fourier transforms we have

$$2\pi D = \int_{|t|>\pi} |g(t)|^2\, dt + \int_{-\pi}^{\pi} \left| g(t) - \varphi(t) \sum_{k=0}^{N} a_k^{(N)}(1 - e^{it})^k \right|^2 dt.$$

Clearly, then, D cannot be made arbitrarily small unless $g(t) = 0$ almost everywhere outside $(-\pi, \pi)$; and if this condition is satisfied, D reduces to the same form which it had in the discrete case—see (19). Thus the problem of mean-square approximation in the continuous case reduces, if it can be solved at all, to the corresponding problem in the discrete case.

11. Representation by a series. We consider the representation of a given $f(x)$ by the B-series with the classical coefficients (5), but with mean-square convergence of the series. Here we assume that $f(x) \geqq 0$, $f(x) = 0$ for $x = -1, -2, \cdots$, and $\sum_{x=0}^{\infty} [f(x)]^2 < \infty$, ask whether we can have

$$\lim_{n\to\infty} \sum_{x=-\infty}^{\infty} \left| f(x) - \sum_{k=0}^{n} a_k \Delta^k \theta(x) \right|^2 = 0, \tag{26}$$

where here the a_k do not depend on n (but are not, in principle, required to have the form (5)). From our previous discussion this is equivalent to

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} \left| g(t) - \varphi(t) \sum_{k=0}^{n} a_k(1 - e^{it})^k \right|^2 dt = 0,$$

and this implies that

$$\lim_{n\to\infty} |a_n|^2 \int_{-\pi}^{\pi} |\varphi(t)|^2 |1 - e^{it}|^{2n} dt = 0.$$

From this it follows easily that

$$\sum_{n=0}^{\infty} a_n(1 - e^{it})^n$$

converges for $|t| < \pi$, or in other words that

$$H(z) = \sum_{n=0}^{\infty} a_n(1 - z)^n$$

converges on $|z| = 1$ except perhaps for $z = -1$, and hence converges in $|1 - z| < 2$ By analytic continuation it is easy to identify $H(z)$ with $F(z)\Phi(z)$, where for $|z| < 1$,

$$F(z) = \sum_{n=0}^{\infty} f(n)z^n, \qquad \Phi(z) = \sum_{n=0}^{\infty} \theta(n)z^n = e^{\lambda(1-z)}.$$

Since $1/\Phi(z)$ has no singular points, $F(z)$ is analytic in $|1 - z| < 2$ and hence in particular in $0 \leqq x < 3$; since $F(z)$ is a power series with nonnegative coefficients, it has a singular point at the positive real point on its circle of convergence [30, p. 214], and so it must be analytic at least in $|z| < 3$. This gives the restriction $\limsup_{n\to\infty} |f(n)|^{1/n} \leqq \frac{1}{3}$. Nevertheless, as we know, $f(x)$ is represented in mean-square by a sequence of sums of terms $a_k^{(N)}\Delta^k\theta(x)$ even if we assume only that $\Sigma |f(n)|^2$ converges.

In the continuous case, if $f(x) \geqq 0$ and we have

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left| f(x) - \sum_{k=0}^{n} a_k\Delta^k\theta(x) \right|^2 dx = 0, \tag{27}$$

we must have $g(x) = 0$ almost everywhere outside $(-\pi, \pi)$ and then, as we saw previously, (26) holds also. Now since $f(x) \geqq 0$, $g(t)$ has derivatives of all orders if it has derivatives of all orders at $t = 0$ [29, p. 90] and it is easily seen from this that $g(t)$ is analytic for all real t if it is analytic at $t = 0$. Now on the one hand, unless $f(x) \equiv 0$, $g(t)$ cannot be analytic for all real t if (as we are supposing) $g(t)$ vanishes outside $(-\pi, \pi)$. On the other hand, $H(e^{it}) = g(t)/\varphi(t)$ for real values of t close to 0 and so, if t is regarded as a complex variable, for complex values of t near 0. Since $1/\varphi(t)$ is analytic everywhere, $g(t)$ is analytic at $t = 0$. From this contradiction we infer that a nonnegative $f(x)$ can never be represented in the form (27), although it may perfectly well be represented by

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left| f(x) - \sum_{k=0}^{n} a_k^{(n)}\Delta^k\theta(x) \right|^2 dx = 0.$$

REFERENCES

I. Charlier series

[1] A C. AITKEN, *Statistical Mathematics*, Oliver and Boyd, London, 1939, pp 58–59, 66–67, 76–79.

[1a] W ANDERSSON, "Short notes on Charlier's method for expansion of frequency functions in series," *Skandinavisk Aktuar. tids.*, Vol. 27 (1944), pp 16–31.

[2] L. A. AROIAN, "The type B Gram-Charlier series," *Annals of Math. Stat.*, Vol. 8 (1937), pp 183–192.

[3] R P. BOAS, JR , "The Charlier B-series," *Am. Math. Soc. Trans.* (to appear)

[4] C. V L CHARLIER, "Über das Fehlergesetz," *Ark Mat. Astr. Fys.*, Vol. 2, no. 8 (1905)

[5] C. V. L CHARLIER, "Die zweite Form des Fehlergesetzes," *ibid* , no 15 (1905).

[6] C. V L. CHARLIER, "Über die Darstellung willkurlicher Funktionen," *ibid.*, no. 20 (1905)

[7] C V. L CHARLIER, "Researches into the theory of probability," *Lund Universitets Årsskrift N F* , Afd 2, Vol. 1, no. 5 (1906).

[8] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 36, Cambridge University Press, 1937.

[9] W. P ELDERTON, *Frequency Curves and Correlation*, 3d ed., Cambridge University Press, 1938, pp. 131–132

[10] A FISHER, *An Elementary Treatise on Frequency Curves and their Application in the Analysis of Death Curves and Life Tables*, Macmillan, New York, 1922

[11] A FISHER, *The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods*, Vol. 1, 2d ed , Macmillan, New York, 1922, pp. 271–279

[12] M. JACOB, "Über die Charlier'sche B-Reihe," *Skandinavisk Aktuar tids*, Vol. 15 (1932), pp. 286–291

[13] C JORDAN, "Sur la probabilité des épreuves répétées, le théorème de Bernoulli et son inversion," *Bull. Soc. Math. France*, Vol. 54 (1926), pp 101–137

[14] N R. JØRGENSEN, "Note sur la fonction de répartition de type B de M. Charlier," *Ark. Mat. Astr. Fys* , Vol. 10, no 15 (1915).

[15] N. R. JØRGENSEN, "Undersøgelser over Frequensflader og Korrelation," *Copenhagen thesis*, 1916.

[16] M. G. KENDALL, *The Advanced Theory of Statistics*, Vol. 1, 3d ed , Griffin, London, 1947, pp 154–156.

[17] S. KULLBACK, "On the Charlier type B series," *Annals of Math Stat.*, Vol. 18 (1947), pp 574–581, Vol. 19 (1948), p. 127

[18] R VON MISES, *Vorlesungen aus dem Gebiete der angewandten Mathematik*, Vol 1, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Deuticke, Leipzig and Vienna, 1931, pp. 265–269.

[19] N. OBRECHKOFF, *Sur la loi de Poisson, la série de Charlier et les équations linéaires aux différences finies de premier ordre à coefficients constants*, Actual. Sci. Ind., no. 740, Hermann et Cie , Paris (1938), pp. 35–64.

[20] H POLLACZEK-GEIRINGER, "Die Charlier'sche Entwicklung willkurlicher Verteilungen," *Skandinavisk Aktuar. tids.*, Vol. 11 (1928), pp 98–111.

[21] H POLLACZEK-GEIRINGER, "Über die Poissonsche Verteilung und die Entwicklung willkurlicher Verteilungen," *Zeits. f. Angew Math. und Mech* , Vol. 8 (1928), pp. 292–309

[22] H L. RIETZ, *Mathematical Statistics*, Carus Mathematical Monographs, no. 3, Chicago, 1927, pp. 60–68, 170–177

[23] E SCHMIDT, *Über die Charliersche Entwicklung einer "arithmetischen Verteilung" nach den sukzessiven Differenzen der Poissonschen asymptotischen Darstellungsfunktion*

für die Wahrscheinlichkeit seltener Ereignisse, Sitzungsberichte der Preussischen Akademie der Wissenschaften, Phys.-Math. Kl. 1928, p. 148.

[24] E. Schmidt, "Über die Charlier-Jordansche Entwicklung einer willkürlichen Funktion nach der Poissonschen Funktion und ihren Ableitung," *Zeits. f. Angew. Math. und Mech.*, Vol. 14 (1933), pp. 139–142.

[25] H. L. Selberg, "Über die Darstellung willkürlicher Funktionen durch Charliersche Differenzreihen," *Skandinavisk Aktuar. tids.*, Vol. 25 (1942), pp. 228–246.

[26] H. L. Selberg, "Über die Darstellung der Dichtefunktion einer Verteilung durch eine Charliersche B-Reihe," *Archiv for Mathematik og Naturvidenskab*, Vol. 46 (1943), pp. 127–138.

[27] J. F. Steffensen, *Some Recent Researches in the Theory of Statistics and Actuarial Science*, Cambridge University Press, 1930, pp. 35–48.

[27a] J. F. Steffensen, "Note sur la fonction de type B de M. Charlier," *Svenska Aktuar. Tids.*, Vol. 3 (1916), pp. 226–228.

[28] J. V. Uspensky, "On Charles Jordan's series for probability," *Annals of Math.* Series 2, Vol. 32 (1931), pp. 306–312.

II Other references

[29] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[30] E. C. Titchmarsh, *The Theory of Functions*, Oxford, 1932.

[31] A. Zygmund, *Trigonometrical Series*, Warszawa-Lwów, 1935.

[32] *Table of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments*, prepared by the Mathematical Tables Project, National Bureau of Standards, 2d ed., Columbia University Press, New York, 1947.

HEURISTIC APPROACH TO THE KOLMOGOROV-SMIRNOV THEOREMS[1]

By J. L. Doob

University of Illinois

1. Introduction and summary. Asymptotic theorems on the difference between the (empirical) distribution function calculated from a sample and the true distribution function governing the sampling process are well known. Simple proofs of an elementary nature have been obtained for the basic theorems of Komogorov[2] and Smirnov[3] by Feller,[4] but even these proofs conceal to some extent, in their emphasis on elementary methodology, the naturalness of the results (qualitatively at least), and their mutual relations. Feller suggested that the author publish his own approach (which had also been used by Kac), which does not have these disadvantages, although rather deep analysis would be necessary for its rigorous justification. The approach is therefore presented (at one critical point) as heuristic reasoning which leads to results in investigations of this kind, even though the easiest proofs may use entirely different methods.

No calculations are required to obtain the qualitative results, that is the existence of limiting distributions for large samples of various measures of the discrepancy between empirical and true distribution functions. The numerical evaluation of these limiting distributions requires certain results concerning the Brownian movement stochastic process and its relation to other Gaussian processes which will be derived in the Appendix.

2. The problem. Let $x_1, x_2, \cdots$ be mutually independent random variables with a common distribution function $F(\lambda)$,

$$F(\lambda) = Pr\{x_j \leq \lambda\}.$$

In statistical language $x_1, \quad , x_n$ form a sample of n drawn from the distribution with distribution function $F(\lambda)$ Let $\nu_n(\lambda)$ be the number of these x_j's which are $\leq \lambda$ According to the strong law of large numbers, for each λ

$$\lim_{n\to\infty} \frac{\nu_n(\lambda)}{n} = F(\lambda) \tag{2.1}$$

with probability 1. For fixed n $\nu_n(\lambda)/n$ is itself a distribution function (which depends on the sample values $x_1, \cdots, x_n$) the *empirical distribution function*, and an elaboration of the argument which led to (2.1) shows that (2.1) is true

[1] Research connected with a probability project at Cornell University under an ONR contract

[2] *Inst. Ital. Atti, Giorn*, Vol 4 (1933), pp. 83-91.

[3] *Rec Math. (Matematiceskii Sbornik)*, N.S 6, Vol 48 (1939), pp 3-26, *Bull Math. Univ Moscou*, Vol 2 (1939), fasc 2

[4] *Annals of Math. Stat*, Vol 19 (1948), pp 177-189

uniformly in λ, with probability 1; that is if

$$D_n = \underset{-\infty<\lambda<\infty}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - F(\lambda) \right|, \tag{2.2}$$

then D_n is a random variable and

$$\lim_{n\to\infty} D_n = 0$$

with probability 1.[5] This result would be of limited practical statistical importance except that the distribution of D_n does not depend on the distribution function $F(\lambda)$ if $F(\lambda)$ is continuous. In fact in that case the random variables $F(x_1)$, $F(x_2)$, $\cdots$ are mutually independent and each is uniformly distributed in the interval (0, 1); if $\hat{\nu}_n(\lambda)$ is the number of $F(x_j)$'s $\leq \lambda$, for $j \leq n$,

$$\underset{0\leq\mu\leq 1}{\text{L.U.B.}} \left| \frac{\hat{\nu}_n(\mu)}{n} - \mu \right| = \underset{-\infty<\lambda<\infty}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - F(\lambda) \right|.$$

Thus it is no restriction, replacing x_j by $F(x_j)$ if necessary, in finding the distribution of D_n to assume that $F(\lambda) = \lambda$ for $0 \leq \lambda \leq 1$, and

$$D_n = \underset{0\leq\lambda\leq 1}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - \lambda \right|. \tag{2.2'}$$

The results will hold for D_n defined by (2.2) for any continuous $F(\lambda)$. We shall also consider D_n^+ and D_n^-, defined by

$$\begin{aligned} D_n^+ &= \underset{0\leq\lambda\leq 1}{\text{L.U.B.}} \left[\frac{\nu_n(\lambda)}{n} - \lambda \right], \\ D_n^- &= -\underset{0\leq\lambda\leq 1}{\text{G.L.B.}} \left[\frac{\nu_n(\lambda)}{n} - \lambda \right], \end{aligned} \tag{2.3}$$

and again the results will hold (with the obvious definitions of D_n^+ and D_n^- in the general case) for every continous $F(\lambda)$.

The problem is to find the limiting distributions of (properly normalized) D_n, D_n^+, D_n^- when $n \to \infty$.

3. Derivation of the Kolmogorov and Smirnov theorems. Define

$$x_n(t) = n^{\frac{1}{2}} \left(\frac{\nu_n(t)}{n} - t \right), \qquad 0 \leq t \leq 1.$$

Since $\nu_n(0) = 0$ with probability 1 and $\nu_n(t) - \nu_n(s)$ is the number of successes in n independent trials, with probability $t - s$ of success in each trial, $\nu_n(t) - \nu_n(s)$ has expectation $n(t-s)$ and variance $n(t-s)[1-(t-s)]$. Hence

$$\begin{aligned} E\{x_n(t)\} &= 0, & 0 \leq t \leq 1; \\ E\{[x_n(t) - x_n(s)]^2\} &= (t-s)[1-(t-s)], & 0 \leq s \leq t \leq 1. \end{aligned} \tag{3.1}$$

[5] Cf. M. Fréchet, *Généralités sur les probabilités. Variables aléatoires*, Paris, 1937, pp. 260-261.

Now let $\{x(t)\}$ be a one parameter family of random variables, $0 \leq t \leq 1$ with the following properties:

(a) for each j if $0 \leq t_1 < \cdots < t_j \leq 1$ the j-variate distribution of the random variables $x(t_1), \cdots, x(t_j)$ is Gaussian;

(b) (3.1) holds, that is

$$E\{x(t)\} = 0, \qquad 0 \leq t \leq 1; \tag{3.1'}$$

$$E\{[x(t) - x(s)]^2\} = (t - s)\,[1 - (t - s)], \qquad 0 \leq s \leq t \leq 1.$$

(c) $Pr\{x(0) = 0\} = 1$.

According to the central limit theorem, the j variate distribution of $x_n(t_1), \cdots, x_n(t_j)$ is asymptotically that of $x(t_1), \cdots, x(t_j)$; in fact the normalizing factor $n^{\frac{1}{2}}$ in the definition of $x_n(t)$ and the choice of means and variances in (3.1') were made precisely to bring this about. As far as first and second moments are concerned the $x_n(t)$ and $x(t)$ processes are identical; when $n \to \infty$ the distributions, or at least the j variate ones mentioned, become identical also.

We shall assume, until a contradiction frustrates our devotion to heuristic reasoning, that *in calculating asymptotic $x_n(t)$ process distributions when $n \to \infty$ we may simply replace the $x_n(t)$ processes by the $x(t)$ process.* It is clear that this cannot be done in all possible situations, but let the reader who has never used this sort of reasoning exhibit the first counter example

The $x(t)$ process has continuous sample functions (cf. Appendix). Define

$$D = \operatorname*{Max}_{0 \leq t \leq 1} |x(t)|,$$

$$D^+ = \operatorname*{Max}_{0 \leq t \leq 1} x(t),$$

$$D^- = -\operatorname*{Min}_{0 \leq t \leq 1} x(t).$$

Then in accordance with our substitution principle $n^{\frac{1}{2}}D_n$, $n^{\frac{1}{2}}D_n^+$, $n^{\frac{1}{2}}D_n^-$ have as n becomes infinite the distributions of D, D^+, D^- respectively. (The latter two are the same because the $-x(t)$ process is stochastically identical with the $x(t)$ process.) Thus these simple qualitative considerations have led to the existence of the limiting distributions derived and evaluated by Kolmogorov, who proved:

THEOREM[6] (Kolmogorov)

$$\lim_{n \to \infty} Pr\{n^{\frac{1}{2}}D_n \geq \lambda\} = 2 \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2}; \tag{3.2}$$

$$\lim_{n \to \infty} Pr\{n^{\frac{1}{2}}D_n^+ \geq \lambda\} = \lim_{n \to \infty} Pr\{nD_n^- \geq \lambda\} = e^{-2\lambda^2}. \tag{3.3}$$

To complete our treatment we shall prove in the Appendix that

$$Pr\{D \geq \lambda\} = 2 \sum_1^\infty (-)^{m+1} e^{-2m^2\lambda^2}, \tag{3.2'}$$

---

[6] In Feller's paper (*loc. cit.*, p. 178, equation (1.4)) the factor 2 in the exponent was omitted by the printer. The same misprint occurs in Smirnov's table of the values of the series in our (3.2), *Annals of Math Stat*, Vol. 19 (1948), pp. 279–281.

$$(3.3') \qquad Pr\{D^+ \geq \lambda\} = Pr\{D^- \geq \lambda\} = e^{-2\lambda^2},$$

so that in fact the above considerations have led not only to the existence but to the evaluation of the asymptotic distributions. (Actually we shall prove somewhat more general results about the $x(t)$ process.)

So much for the Kolmogorov theorems Smirnov obtained results (also independent of the given continuous distribution function $F(\lambda)$) of a somewhat different nature. Let $x_1^*, x_2^*, \cdots$ be mutually independent random variables with the same individual distributions as the x_j's, that is each distributed uniformly in the interval (0, 1); define $\nu_n^*(\lambda)$ as the number of the first n x_j's which are $\leq \lambda$ Smirnov considered the difference between empirical distribution functions,

$$D_{mn} = \text{L.U.B.}_{0 \leq \lambda \leq 1} \left| \frac{\nu_m(\lambda)}{m} - \frac{\nu_n^*(\lambda)}{n} \right|,$$

as well as D_{mn}^+ and D_{mn}^- defined in the obvious way. To avoid stressing the obvious we consider only the D_{mn}.

THEOREM (Smirnov). *If* $m, n \to \infty$ *in such a way that* $\frac{m}{n} \to r$, *and if* $N = mn/(m + n)$,

$$(3.4) \qquad \lim_{n \to \infty} Pr\{N^{\frac{1}{2}} D_{mn} \geq \lambda\} = 2 \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2}.$$

To derive this result define an $x^*(t)$ process stochastically identical with the $x(t)$ process but independent of it. Then if $x_n^*(t)$ is defined by

$$x_n^*(t) = n^{\frac{1}{2}} \left(\frac{\nu_n^*(t)}{n} - t \right),$$

we identify, in accordance with our heuristic principle the process with variables

$$\{x(t) - r^{\frac{1}{2}} x^*(t)\}$$

with the one with variables

$$\left\{ x_m(t) - \left(\frac{m}{n} \right)^{1/2} x_n^*(t) \right\}.$$

Doing this leads to the fact that the distribution of

$$(N)^{1/2} D_{mn} = \left(\frac{n}{m + n} \right)^{1/2} \text{L.U.B.}_{0 \leq t \leq 1} \left| x_m(t) - \left(\frac{m}{n} \right)^{1/2} x_n^*(t) \right|$$

converges to that of

$$\left(\frac{1}{1 + r} \right)^{1/2} \text{Max}_{0 \leq t \leq 1} | x(t) - (r)^{1/2} x^*(t) |.$$

Now the $x(t)$ process and the process with variables

$$\left\{\frac{x(t) - (r)^{1/2}x^*(t)}{(1+r)^{1/2}}\right\}$$

are stochastically identical. Hence we are led to the conclusion that the distribution of $(N)^{1/2}D_{mn}$ converges to that of D, and this is Smirnov's theorem, stated above (The method we use does not seem applicable to Smirnov's deeper theorems on the number of intersections between empirical and true distribution curves or between pairs of empirical distribution curves.)

APPENDIX

4. The Brownian movement process. Consider any Gaussian stochastic process, with random variables $\{x(t)\}$ where t varies in some interval. That is, we assume that for each t in the interval $x(t)$ is a random variable and that for any $j \geq 1$ if $t_1 < \cdots < t_j$ are in the interval the j variate distribution of $x(t_1), \cdots, x(t_j)$ is Gaussian. In the following we shall always assume that $E\{x(t)\} \equiv 0$. Then the process is determined stochastically by the covariance function

$$r(s, t) = E\{x(s)x(t)\}.$$

In particular, if the range of parameter is the interval $[0, \infty)$ and if

$$r(s, t) = \sigma^2 \operatorname{Min}(s, t), \qquad 0 \leq s, t < \infty,$$

the process is called the Brownian movement process, or sometimes the Wiener process; σ is a positive constant. When considering this process we shall write $\zeta(t)$ instead of $x(t)$. For the $\zeta(t)$ process

$$Pr\{\zeta(0) = 0\} = 1,$$
$$E\{[\zeta(t) - \zeta(s)]^2\} = \sigma^2 \mid t - s \mid,$$

and if $0 \leq s_1 < t_1 < s_2 < t_2$ the increments $x(t_1) - x(s_1)$ and $x(t_2) - x(s_2)$ are mutually independent. We shall use the following properties of this process, of which the first two are well known.

(a) The sample functions are everywhere continuous with probability 1. In the following we can therefore write as if all the sample curves were continuous.

(b) For fixed s

$$(4.1) \qquad Pr\{\operatorname*{Max}_{0<t\leq T}[\zeta(s+t) - \zeta(s)] \geq \lambda\} = 2Pr\{\zeta(s+T) - \zeta(s) \geq \lambda\}.^7$$

(Note that the use of a general initial value s, rather than 0, has not added to the generality and we drop this affectation below.)

(c) If $a \geq 0$, $b > 0$, $\alpha \geq 0$, $\beta > 0$, then

$$(4.2) \qquad Pr\{\operatorname*{L.U.B.}_{0\leq t<\infty}[\zeta(t) - (at + b)] \geq 0\} = e^{-2ab/\sigma^2},$$

---

[7] Due to Bachelier, cf the proof by P Lévy, *Comp Math*, Vol 7 (1939), p 293 One way to prove (*a*) is to prove (4 1) first, with L.U B instead of Max, and then use it to calculate the probabilities relevant to (*a*)

$$(4.3)\quad Pr\{\underset{0\le t<\infty}{\text{L.U.B.}}\,[\zeta(t)-(at+b)]\ge 0 \quad or \quad \text{G.L.B.}\,[\zeta(t)+\alpha t+\beta]\le 0\},$$

$$=\sum_{m=1}^{\infty}\{e^{-2[m^2ab+(m-1)^2\alpha\beta+m(m-1)(a\beta+\alpha b)]}$$

$$+e^{-2[(m-1)^2ab+m^2\alpha\beta+m(m-1)(a\beta+\alpha b)]}$$

$$-e^{-2[m^2(ab+\alpha\beta)+m(m-1)\alpha\beta+m(m+1)ab]}$$

$$-e^{-2[m^2(ab+\alpha\beta)+m(m+1)a\beta+m(m-1)\alpha b]}\};$$

in particular ($\alpha=a$, $\beta=b$)

$$(4.3')\qquad Pr\left\{\underset{0\le t<\infty}{\text{L.U.B.}}\,\frac{|\zeta(t)|}{at+b}\ge 1\right\}=2\sum_{1}^{\infty}(-1)^{m+1}e^{-2m^2ab}.$$

The probability in (4.2) is the probability that a $\zeta(t)$ sample curve will ever reach the line with slope a and ordinate intercept b; the probability in (4.3) is the probability that a sample curve will ever reach either of the indicated halflines, one above and one below the t axis. Since the right hand sides are continuous functions of a, b, α, β we could write >0 instead of ≥ 0 and <0 instead of ≤ 0 on the left, so that these probabilities are also the probabilities that a sample curve will ever rise above the indicated line or leave the indicated angle.

It will be convenient to describe a line by its slope and ordinate intercept; the line $[u, v]$ is the line with slope u and ordinate intercept v. We shall take $\sigma=1$ in the proof; this is no essential restriction since $\zeta(t)/\sigma$ is the random variable of a process of the same type whose σ is 1.

To prove (4.2) let $\varphi(a, b)$ be the probability on the left, the probability that a sample curve will reach the line $[a, b]$. If $b=b_1+b_2$, $b_i>0$, a sample curve which is to reach $[a, b]$ must first reach $[a, b_1]$ and then move up to meet a line with slope a, b_2 units above the first meeting with $[a, b_1]$. Then

$$\varphi(a, b_1+b_2)=\varphi(a, b_1)\,\varphi(a, b_2).$$

Now $\varphi(a, b)\ge Pr\{\zeta(1)\ge a+b\}>0$ and $\varphi(a, b)$ is monotone non-increasing in b, for fixed a. The only solution of the functional equation with these properties is

$$\varphi(a, b)=e^{-\psi(a)b}.$$

Now $\varphi(a, b)$ is the probability of reaching $[0, b]$ at some first time s and then going on to the line $[a, b]$ which from the vantage point of the first common point $(s, \zeta(s))$ is the line $[a, as]$. In other words, using (4.1)

$$e^{-\psi(a)b}=-\int_0^{\infty}e^{-\psi(a)as}\,d_s\,Pr\{\underset{0\le t\le s}{\text{Max}}\,\zeta(t)\ge b\}$$

$$=\int_0^{\infty}e^{-\psi(a)as}\,\frac{be^{-(b^2)/2s}}{(2\pi)^{1/2}s^{3/2}}\,ds$$

$$= \frac{2}{\pi^2}\int_0^\infty \exp\left[-s^2 - \frac{b^2 a\psi(a)}{2s^2}\right] ds$$
$$= e^{-b(2a\psi(a))^{1/2}},$$

from which it follows that $\psi(a) = 2a$, and this yields (4.2).

To prove (4.3) we consider first the following general problem: Let $[u_1, v_1]$, $[u_2, v_2]$, $\cdots$, $u_j \geq 0, v_j \geq 0$ be a sequence of lines; let $t = t_1$ be the first value of t, if any, at which a sample curve meets $[u_1, v_1]$; if t_1 is defined for a sample curve let t_2 be the first value of $t > t_1$, if any at which the curve meets $[-u_2, -v_2]$; if t_2 is defined for a sample curve, let t_3 be the first value of $t > t_2$, if any, at which the curve meets $[u_3, v_3]$, and so on. Let π_n be the probability that there is a point t_n, in other words the probability that a sample curve meets the lines $[u_1, v_1]$, $[-u_2, -v_2]$, $\cdots$ $[(-1)^{n+1}u_n, (-1)^{n+1}v_n]$ in at least n successive points. We write

$$\pi_n = \pi_n(u_1, v_1, \cdots, u_n, v_n).$$

In particular, according to (4.2)

$$\pi_1(u_1, v_1) = e^{-2u_1v_1}. \tag{4.4}$$

To evaluate π_n, let Q be the point $(t_{n-1}, \zeta(t_{n-1}))$ on the sample curve, and suppose for definiteness that n is even. Starting at Q, if there is a t_n, the curve must finally reach $[-u_n, -v_n]$, that is it must go to a line of slope $-u_n$, which is $u_{n-1}t_{n-1} + v_{n-1} + u_nt_{n-1} + v_n$ units vertically below its initial position Q when $t = t_{n-1}$. According to (4.2) the probability of doing this is

$$e^{-2u_n(u_{n-1}t_{n-1}+v_{n-1}+u_nt_{n-1}+v_n)}.$$

Now we replace the line $[-u_n, -v_n]$ by a line which depends on t_{n-1} but which leaves this probability unchanged; the new line has slope $-(u_{n-1} + u_n)$ and is

$$h = \frac{u_n}{u_{n-1} + u_n}(u_{n-1}t_{n-1} + v_{n-1} + u_nt_{n-1} + v_n)$$

units below Q when $t = t_{n-1}$. Finally we reflect this new line in the line parallel to the t axis through Q. These two changes do not affect the probability we are discussing because the changes of $\zeta(t)$ after t_{n-1} are independent of the changes before and have symmetric distributions. The final line has slope $u_{n-1} + v_{n-1}$ and is h units above Q when $t = t_{n-1}$; it is the line

$$\left[u_{n-1} + u_n, \quad \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n}\right]$$

which does not depend on t_{n-1}. This line lies above $[u_{n-1}, v_{n-1}]$ in the first quadrant, so that if a sample curve reaches it the curve must also intersect $[u_{n-1}, v_{n-1}]$. We have thus proved that

$$\pi_n(u_1, v_1; \cdots; u_n, v_n) \tag{4.5}$$
$$= \pi_{n-1}\left(u_1, v_1; \cdots; u_{n-2}, v_{n-2}; u_{n-1} + u_n, \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n}\right).$$

The fundamental identity (4.5) makes it possible to reduce the evaluation of π_n to π_1 in $n-1$ steps, π_1 is evaluated in (4.4). Thus successive meetings with n lines have been reduced to a meeting with a single line. As a first example suppose

$$u_1 = \cdots = u_n = u, \qquad v_1 = \cdots = v_n = v.$$

Then we have

$$\begin{aligned}\pi_n(u, v, \cdots ; u, v) &= \pi_{n-1}(u, v, \cdots ; 2u, 2v) = \cdots \\ &= \pi_1(nu, nv),\end{aligned}$$

so that

$$\pi_n(u, v; \cdots ; u, v) = e^{-2n^2uv} \tag{4.6}$$

More generally suppose

$$u_1 = u_3 = \cdots = a, \qquad v_1 = v_3 = \cdots = b,$$

$$u_2 = u_4 = \cdots = \alpha, \qquad v_2 = v_4 = \cdots = \beta.$$

Then we show that for suitably chosen $C_j^{(n)}$'s we have according as n is even or odd

$$\pi_n(a, b; \cdots ; \alpha, \beta) = \pi_1\left[\frac{n}{2}(a+\alpha), \frac{C_1^{(n)}ab + C_2^{(n)}\alpha\beta + C_3^{(n)}a\beta + C_4^{(n)}\alpha b}{\frac{n}{2}(a+\alpha)}\right];$$

$$\begin{aligned}&\pi_n(a, b; \cdots ; a, b) \\ &= \pi_1\left[\frac{n+1}{2}a + \frac{n-1}{2}\alpha, \frac{C_1^{(n)}ab + C_2^{(n)}\alpha\beta + C_3^{(n)}a\beta + C_4^{(n)}\alpha b}{\frac{n+1}{2}a + \frac{n+1\alpha}{2}}\right].\end{aligned} \tag{4.7}$$

For $n = 1$ this form is correct with

$$C_1^{(1)} = 1, \qquad C_2^{(1)} = C_3^{(1)} = C_4^{(1)} = 0$$

If now n is even and if the equations are true for n,

$$\pi_{n+1}(a, b; \cdots; a, b) = \pi_2\left(a, b; \frac{n}{2}(\alpha + a), \frac{C_1^{(n)}\alpha\beta + C_2^{(n)}ab + C_3^{(n)}\alpha b + C_4^{(n)}a\beta}{\frac{n}{2}\alpha + a}\right)$$

$$= \pi_1\left(\frac{n+2}{2}a + \frac{n}{2}\alpha, \frac{ab + C_1^{(n)} + C_2^{(n)}ab + C_3^{(n)}\alpha b + C_4^{(n)}a\beta + n(\alpha + a)b}{\frac{n+2}{2}a + \frac{n}{2}\alpha}\right),$$

and comparing this with (4.7) we find that

$$C_1^{(n+1)} = C_2^{(n)} + n + 1,$$

$$\begin{aligned} C_2^{(n+1)} &= C_1^{(n)}, \\ C_3^{(n+1)} &= C_4^{(n)}, \\ C_4^{(n+1)} &= C_3^{(n)} + n, \end{aligned} \qquad (n \text{ even}).$$

If n is odd we find similarly that

$$\begin{aligned} C_1^{(n+1)} &= C_2^{(n)} + n, \\ C_2^{(n+1)} &= C_1^{(n)}, \\ C_3^{(n+1)} &= C_4^{(n)}, \\ C_4^{(n+1)} &= C_3^{(n)} + n + 1. \end{aligned}$$

The solution of these equations is

n *even*	n *odd*
$C_1^{(n)} = \frac{n^2}{4}$	$C_1^{(n)} = \frac{(n+1)^2}{4}$
$C_2^{(n)} = \frac{n^2}{4}$	$C_2^{(n)} = \frac{(n-1)^2}{4}$
$C_3^{(n)} = \frac{n(n-2)}{4}$	$C_3^{(n)} = \frac{n^2-1}{4}$
$C_4^{(n)} = \frac{n(n+2)}{4}$	$C_4^{(n)} = \frac{n^2-1}{4}$

Then

$$\begin{aligned} \pi_n &= e^{-\frac{1}{2}[n^2\alpha b + n2\alpha\beta + n(n-2)\alpha\beta + n(n+2)\alpha b]} && (n \text{ even}), \\ \pi_n &= e^{-\frac{1}{2}[(n+1)^2\alpha b + (n-1)^2\alpha\beta + (n^2-1)\alpha\beta + (n^2-1)\alpha b]} && (n \text{ odd}). \end{aligned} \tag{4.8}$$

We can now prove (4 3). In fact the left side is equal to

$$\pi_1(a, b) + \pi_1(\alpha, \beta) - \pi_2(a, b; \alpha, \beta) - \pi_2(\alpha, \beta; a, b) + \cdots,$$

which gives (4.3), on substituting (4 8) Only (4.3′), which follows from the simple (4.6), is used in the application to the Kolmogorov-Smirnov theorems.

5. Transformations of Gaussian processes to the Brownian movement process. The $\zeta(t)$ process studied in section 4 is so simple that it is important to be able to reduce others to it by elementary changes of variable For example if the covariance function of a Gaussian process has the form

$$r(s, t) = u(s)v(t), \qquad s < t, \tag{5.1}$$

for s, t in some interval, and if the ratio

$$\frac{u(t)}{v(t)} = a(t)$$

is continuous and monotone increasing, with inverse function $a_1(t)$. We define

$$\zeta(t) = \frac{u[a_1(t)]}{v[a_1(t)]}.$$

With this definition the ζ process is Gaussian and since if $s < t$

$$E\{\zeta(s)\zeta(t)\} = \frac{u[a_1(s)]v[a_1(t)]}{v[a_1(s)]v[a_1(t)]} = a[a_1(s)] = s = \text{Min}(s, t),$$

the ζ process is the Brownian movement process with $\sigma = 1$. This transformation from the x to the ζ process is effected by a combination of a change of variable in t and the application of a variable scaling factor. (Conversely, if such a transformation is applied to the Brownian movement process it is trivial to verify that the new covariance function will have the form (5.1). The Gaussian processes with covariance functions of this form are easily seen to be the Gaussian Markov processes)

6. The Gaussian process with r(s, t) = s(1 − t). In section 3 the Kolmogorov-Smirnov theorems were reduced to properties of a Gaussian process with parameter t, $0 \leq t \leq 1$, for which

$$Pr\{x(0) = 0\} = 1;$$

$$E\{x(t)\} = 0;$$

$$E\{[x(t) - x(s)]^2\} = (t - s)[1 - (t - s)], \quad 0 \leq s < t \leq 1.$$

Now these equations imply that

$$E\{x(t)^2\} = t(1 - t), \qquad E\{x(s)^2\} = s(1 - s),$$

and combining the set we find that

$$r(s, t) = E\{x(s)x(t)\} = s(1 - t), \qquad 0 \leq s < t \leq 1.$$

This covariance function has the form studied in section 5, and using the transformation of that section

$$\zeta(t) = (t + 1)x\left(\frac{t}{t + 1}\right), \qquad 0 \leq t < \infty,$$

defines a Brownian movement process (with $\sigma = 1$). Then if D, D^+, D^- are defined as in section 3, we have from (4.3′)

$$Pr\{D \geq \lambda\} = Pr\left\{\text{L. U. B.}_{0 \leq t < \infty} \left| \frac{\zeta(t)}{t + 1} \right| \geq \lambda\right\} = \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2},$$

and from (4.2)

$$Pr\{D^+ \geq \lambda\} = Pr\{D^- \geq \lambda\} = e^{-2\lambda^2}.$$

This proves (3.2′) and (3.3′). Note that we could go beyond these results, because of our detailed knowledge of the $x(t)$ process. For example we can evaluate

$$\lim_{n\to\infty} Pr\{(n)^{\frac{1}{2}}D_n^- \leq \lambda_1\,, \qquad (n)^{\frac{1}{2}}D_n^+ \leq \lambda_2\}.$$

If $\lambda_1 = \lambda_2 = \lambda$ the probability is the probability that $(n)^{1/2}D_n \leq \lambda$ which we have already treated. In general it is, in the limit,

$$\begin{aligned}
&Pr\{\operatorname*{Min}_{0\leqq t\leqq 1} x(t) \geq -\lambda_1\,, \operatorname*{Max}_{0\leq t\leq 1} x(t) \leq \lambda_2\} \\
&\quad = Pr\left\{\operatorname*{G.L.B.}_{0\leqq t\leqq\infty} \frac{\zeta(t)}{t+1} \geq -\lambda_1 \operatorname*{L.U.B.}_{0\leq t<\infty} \frac{\zeta(t)}{t+1} \leq \lambda_2\right\} \\
&\quad = 1 - \sum_{m=1}^{\infty} \{e^{-2[m^2\lambda_2^2+(m-1)^2\lambda_1^2+2m(m-1)\lambda_1\lambda_2]} + e^{-2[(m-1)^2\lambda_2^2+m^2\lambda_1^2+2m(m-1)\lambda_1\lambda_2]} \\
&\qquad - e^{-2[m^2(\lambda_1^2+\lambda_2^2)+m(m-1)\lambda_1\lambda_2+m(m+1)\lambda_1\lambda_2]} - e^{-2[m^2(\lambda_1^2+\lambda_2^2)+m(m+1)\lambda_1\lambda_2+m(m-1)\lambda_1\lambda_2]}\} \\
&\quad = 1 - \sum_{m=1}^{\infty} \{e^{-2[m\lambda_2+(m-1)\lambda_1]^2} + e^{-2[(m-1)\lambda_2+m\lambda_1]^2} - 2e^{-2m^2(\lambda_1+\lambda_2)^2}\}
\end{aligned}$$

obtained by setting $a = b = \lambda_2$, $\alpha = \beta = \lambda_1$ in (4.3).

PEARSONIAN CORRELATION COEFFICIENTS ASSOCIATED WITH LEAST SQUARES THEORY

By Paul S. Dwyer

University of Michigan

1. Introduction and summary. It is well known that the zero-order correlation between the predicted value of a variable and the observed value of the variable is the multiple correlation. It is also well known that the zero-order correlation between the residuals for two different variables, when the prediction is from a common set of variables, is the partial correlation. These considerations naturally lead to a systematic investigation of all the zero-order correlations involving the various variables associated with least squares theory. Such an investigation is the purpose of this paper.

As a result of this study it appears that other zero-order correlations include the multiple alienation coefficient, the part correlation coefficient, and certain other coefficients which, as far as I am aware, have not been previously defined.

The paper first examines the case of a single predicted variable and then continues with the case in which two or more variables are predicted simultaneously. The paper includes (1) a theoretical development of the different coefficients and the relations between them, (2) the expression of the formulas in determinantal form, (3) a matrix presentation of the material, and (4) an outline of the calculational techniques—with illustrations.

It should be made clear at the start that this paper deals with populations (finite or infinite) and not with samples from those populations. The sampling distribution of each of the new correlation coefficients defined in this paper might well become the subject of a later investigation, but first we need to know what these correlation coefficients are.

2. The case of the single predicted variable. Notation, definitions, and basic properties. We suppose that a population consists of N individuals with values $X_{1j}, X_{2j}, \cdots, X_{kj}, Y_j$ for the variables $X_1, X_2, \cdots, X_k, Y$ and that Y is linearly predicted from the X_i by the formula

$$E = Y - \alpha_0 - \alpha_1 X_1 - \alpha_2 X_2 - \cdots - \alpha_k X_k = Y - \underline{Y} \tag{1}$$

by least squares theory. For the purposes of this paper, we use a concise summation notation, ΣQ, in place of the more formal serial notation $\sum_{i=1}^{N} Q_i$ which is preferable to the frequency notation $\sum_{x=a}^{b} Q_x f_x$ and, in the continuous case, $\int_a^b Q_x f_x dx$ Moreover it is desirable that the scales of X and Y be chosen so as

to facilitate the easy determination of the various formulas. If we let

$$(2) \qquad y_j = \frac{Y_j - \bar{Y}}{\sqrt{N}\sigma_y}; \qquad x_{ij} = \frac{X_{ij} - \bar{X}_i}{\sqrt{N}\sigma_{x_i}}$$

we have $\Sigma x_i^2 = \Sigma y^2 = 1$ with the resulting correlating formula

$$(3) \qquad \rho_{x_i y} = \frac{\Sigma x_i y}{\sqrt{(\Sigma x_i^2)(\Sigma y^2)}} = \Sigma x_i y \quad \text{and} \quad \rho_{x_i x_j} = \Sigma x_i x_j.$$

The transformations (2) when applied to (1) give

$$(4) \qquad e = \frac{E}{\sqrt{N}\sigma_Y} = y - (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) = y - \underline{y}$$

where the β's are standard regression coefficients and e is defined to be $\frac{E}{\sqrt{N}\sigma_Y}$. It is to be noted that the values of x_i, y, e, and $\underline{y}$ are all dimensionless.

The values we wish to correlate are those of X_i, Y, E, $\underline{Y}$ of (1). The zero-order correlations involving these are the same as for x_i, y, e, $\underline{y}$ of (4).

3. Correlations with a single predicted variable. We wish to minimize Σe^2. Differentiating with respect to β_i and equating to zero we get

$$(5) \qquad \Sigma e x_i = 0$$

from which by multiplication by β_i and summation for i,

$$(6) \qquad \Sigma e \underline{y} = 0.$$

It follows that

$$(7) \quad \Sigma e^2 = \Sigma e(y - \underline{y}) = \Sigma e y = \Sigma(y - \underline{y})y = \Sigma y^2 - \Sigma y\underline{y} = 1 - \Sigma y\underline{y}$$
$$= 1 - \Sigma(e + \underline{y})\underline{y} = 1 - \Sigma \underline{y}^2$$

Using (4) and (7), we get

$$\Sigma e^2 = \frac{\Sigma E^2}{N\sigma_Y^2} = \frac{\sigma_E^2}{\sigma_Y^2} = 1 - \Sigma \underline{y}^2$$

so that

$$(8) \qquad \Sigma \underline{y}^2 = \frac{\sigma_Y^2 - \sigma_E^2}{\sigma_Y^2}.$$

This is the conventional definition (from least squares theory) of the multiple correlation coefficient, so

$$(9) \qquad \rho^2_{y:x_1 x_2 \cdots x_k} = \rho^2_{y(x)} = \Sigma \underline{y}^2 = \Sigma y \underline{y}.$$

Application of (9) to (7) gives

$$(10) \qquad \Sigma e^2 = 1 - \rho^2_{y(x)} = \kappa^2_{y(x)} = \kappa^2_{y:x_1 x_2 \cdots x_k}$$

where $\kappa_{y(x)}$ is the multiple alienation coefficient. We now have $\Sigma x_i^2 = 1$, $\Sigma y^2 = 1$, $\Sigma e^2 = \kappa^2_{y(x)}$, and $\Sigma \underline{y}^2 = \rho^2_{y(x)}$, so that we are able to present formulas involving x_i, y, e, $\underline{y}$. We first form the cross products

$$\Sigma xy = \rho_{xy}, \tag{11}$$

$$\Sigma xe = 0, \tag{12}$$

$$\Sigma x\underline{y} = \Sigma x(y + e) = \Sigma xy = \rho_{xy}, \tag{13}$$

$$\Sigma ye = \Sigma y(y - \underline{y}) = \Sigma y^2 - \Sigma y\underline{y} = 1 - \Sigma \underline{y}^2 = \kappa^2_{y(x)}, \tag{14}$$

$$\Sigma y\underline{y} = \Sigma \underline{y}^2 = \rho^2_{y(x)}, \tag{15}$$

$$\Sigma e\underline{y} = 0. \tag{16}$$

We then have

$$\rho_{x_i x_j} = \frac{\Sigma x_i x_j}{\sqrt{(\Sigma x_i^2)(\Sigma x_j^2)}} = \Sigma x_i x_j. \tag{17}$$

$$\rho_{x_i y} = \frac{\Sigma x_i y}{\sqrt{(\Sigma x_i^2)(\Sigma y^2)}} = \Sigma x_i y, \tag{18}$$

$$\rho_{xe} = \frac{\Sigma xe}{\sqrt{(\Sigma x^2)(\Sigma e^2)}} = 0, \tag{19}$$

$$\rho_{x\underline{y}} = \frac{\Sigma x\underline{y}}{\sqrt{(\Sigma x^2)(\Sigma \underline{y}^2)}} = \frac{\Sigma xy}{\rho_{y(x)}} = \frac{\rho_{xy}}{\rho_{y(x)}}. \tag{20}$$

It is interesting to note that this is unity in case $k = 1$ for then $\rho_{x\underline{y}} = \rho_{y(x)}$. Otherwise the absolute value of $\rho_{x\underline{y}}$ is larger than that of ρ_{xy}. For this reason this coefficient might be called the *multiple augmented correlation coefficient.*

$$\rho_{e\underline{y}} = \frac{\Sigma e\underline{y}}{\sqrt{(\Sigma e^2)(\Sigma \underline{y}^2)}} = \frac{\kappa^2_{y(x)}}{\kappa_{y(x)}} = \kappa_{y(x)}. \tag{21}$$

Thus the correlation between y and its residual is the multiple alienation coefficient.

$$\rho_{y\underline{y}} = \frac{\Sigma y\underline{y}}{\sqrt{(\Sigma y^2)(\Sigma \underline{y}^2)}} = \sqrt{\Sigma \underline{y}^2} = \rho_{y(x)}. \tag{22}$$

Thus, as is well known, the zero-order correlation between observed and predicted y is the multiple correlation.

$$\rho_{e\underline{y}} = \frac{\Sigma e\underline{y}}{\sqrt{(\Sigma e^2)(\Sigma \underline{y}^2)}} = 0. \tag{23}$$

4. Notation for the general case. We need to extend the notation and the definitions before examining explicit formulas for the more general case of two (or more) predicted variables. Suppose that Y_1 and Y_2 are the two variables

predicted from the same X's. Then from (4) we write

$$(24)\qquad \begin{aligned} e_i &= \frac{E_i}{\sqrt{N}\sigma_{Y_i}} = y_i - \beta_{i1}x_1 - \beta_{i2}x_2 - \cdots - \beta_{ik}x_k = y_i - \underline{y}_i \\ e_j &= \frac{E_j}{\sqrt{N}\sigma_{Y_j}} = y_j - \beta_{j1}x_1 - \beta_{j2}x_2 - \cdots - \beta_{jk}x_k = y_j - \underline{y}_j . \end{aligned}$$

We then have the two sets of normal equations

$$(25)\qquad \Sigma e_i x = 0 \qquad \Sigma e_j x = 0$$

so that

$$(26)\qquad \begin{aligned} \Sigma e_i \underline{y}_i &= 0 \qquad \Sigma e_j \underline{y}_i = 0 \\ \Sigma e_i \underline{y}_j &= 0 \qquad \Sigma e_j \underline{y}_j = 0. \end{aligned}$$

It follows that

$$(27)\qquad \begin{aligned} \Sigma e_i e_j = \Sigma e_i(y_j - \underline{y}_j) = \Sigma e_i y_j = \Sigma(y_i - \underline{y}_i)y_j = \Sigma y_i y_j - \Sigma \underline{y}_i y_j \\ = \Sigma y_i y_j - \Sigma \underline{y}_i \underline{y}_j = \Sigma y_i y_j - \Sigma \underline{y}_i \underline{y}_j = \rho_{ij} - \Sigma \underline{y}_i \underline{y}_j \end{aligned}$$

if we use the notation that $\rho_{ij} = \rho_{y_i y_j}$.

5. The correlations involving more than one predicted variable. In this case the y's, the e's and the $\underline{y}$'s (as well as the x's) can have more than one variable so that the correlation coefficients we need, in addition to those of section 3, are $\rho_{y_i y_j}$, $\rho_{e_i e_j}$, $\rho_{y_i y_j}$, $\rho_{y_i e_j}$, $\rho_{y_j e_i}$, $\rho_{y_i y_j}$, $\rho_{y_i y_j}$, $\rho_{e_i y_j}$, and $\rho_{y_i e_j}$. We need now only the summed products

$$(28)\qquad \begin{aligned} &\Sigma y_i y_j = \rho_{y_i y_j} = \rho_{ij}, \\ &\Sigma e_i e_j = \rho_{ij} - \Sigma \underline{y}_i \underline{y}_j \qquad \text{as given in (27)}, \end{aligned}$$

$$(29)\qquad \Sigma y_i e_j = \Sigma y_i(y_j - \underline{y}_j) = \Sigma y_i y_j - \Sigma y_i \underline{y}_j = \rho_{ij} - \Sigma \underline{y}_i \underline{y}_j,$$

$$(30)\qquad \Sigma y_i \underline{y}_j = \Sigma \underline{y}_i \underline{y}_j,$$

$$(31)\qquad \Sigma e_i \underline{y}_j = 0.$$

We have then

$$(32)\qquad \rho_{ij} = \frac{\Sigma y_i y_j}{\sqrt{(\Sigma y_i^2)(\Sigma y_j^2)}} = \Sigma y_i y_j,$$

$$(33)\qquad \rho_{e_i e_j} = \frac{\Sigma e_i e_j}{\sqrt{(\Sigma e_i^2)(\Sigma e_j^2)}} = \frac{\rho_{ij} - \Sigma \underline{y}_i \underline{y}_j}{\kappa_{i(x)}\kappa_{j(x)}}.$$

This is the partial correlation coefficient.

$$(34)\qquad \rho_{\underline{y}_i \underline{y}_j} = \frac{\Sigma \underline{y}_i \underline{y}_j}{\sqrt{(\Sigma \underline{y}_i^2)(\Sigma \underline{y}_j^2)}} = \frac{\Sigma \underline{y}_i \underline{y}_j}{\rho_{i(x)}\rho_{j(x)}}.$$

This coefficient appears to be new. Since it is the correlation of predicted values, I suggest that it be called the *predictions correlation coefficient*

$$\rho_{y_i e_j} = \frac{\Sigma y_i c_j}{\sqrt{(\Sigma y_i^2)(\Sigma e_j^2)}} = \frac{\rho_{ij} - \Sigma y_i y_j}{\kappa_j(x)}, \tag{35}$$

$$\rho_{e_i y_j} = \frac{\Sigma e_i y_j}{\sqrt{(\Sigma e_i^2)(\Sigma y_j^2)}} = \frac{\rho_{ij} - \Sigma y_i y_j}{\kappa_i(x)}. \tag{36}$$

The correlations given by (35) and (36) have been defined previously and are known as part correlation coefficients [1; 213,497].

$$\rho_{y_i y_j} = \frac{\Sigma y_i y_j}{\sqrt{(\Sigma y_i^2)(\Sigma y_j^2)}} = \frac{\Sigma y_i y_j}{\rho_j(x)}, \tag{37}$$

$$\rho_{y_i y_j} = \frac{\Sigma y_i y_j}{\sqrt{(\Sigma y_i^2)(\Sigma y_j^2)}} = \frac{\Sigma y_i y_j}{\rho_i(x)}. \tag{38}$$

The correlations of (37) and (38) appear to be new. Each is, in a sense, a generalization of the multiple correlation coefficient since it becomes the multiple correlation coefficient when $i = j$ I suggest that it might be called the *cross multiple correlation coefficient*, since it correlates the actual value of one variable with the predicted value of another.

$$\begin{aligned} \rho_{e_i y_j} &= \frac{\Sigma e_i y_j}{\sqrt{(\Sigma e_i^2)(\Sigma y_j^2)}} = 0, \\ \rho_{y_i e_j} &= \frac{\Sigma y_i e_j}{\sqrt{(\Sigma y_i^2)(\Sigma e_j^2)}} = 0. \end{aligned} \tag{39}$$

A summary of definitions and names of Pearsonian correlation coefficients associated with least squares theory is presented in Table I. No name is proposed when the coefficient is identically zero.

6. Relations between the correlations. Many relations exist between the correlations defined in earlier sections. Some of the more interesting of these are obtained by the elimination of $\Sigma y_i y_j$ from formulas involving this term. Thus from (34), (37), and (38) we get

$$\Sigma y_i y_j = \rho_{y_i y_j}\rho_i(x)\rho_j(x) = \rho_{y_i y_j}\rho_j(x) = \rho_{y_i y_j}\rho_i(x),$$

and from (33), (35), and (36) we get

$$\rho_{ij} - \Sigma y_i y_j = \rho_{e_i e_j}\kappa_i(x)\kappa_j(x) = \rho_{y_i e_j}\kappa_j(x) = \rho_{e_i y_j}\kappa_i(x).$$

We then have

$$\left.\begin{aligned} \rho_{ij} - \rho_{y_i y_j}\rho_i(x)\rho_j(x) \\ \rho_{ij} - \rho_{y_i y_j}\rho_j(x) \\ \rho_{ij} - \rho_{y_i y_j}\rho_i(x) \end{aligned}\right\} = \left\{\begin{aligned} \rho_{e_i e_j}\kappa_i(x)\kappa_j(x) \\ \rho_{y_i e_j}\kappa_j(x) \\ \rho_{e_i y_j}\kappa_i(x) \end{aligned}\right. \tag{40}$$

where the six members may be equated in all possible ways.

Interesting and simple relations can also be obtained by formation of ratios. Thus

$$\text{(41)}\qquad \frac{\rho_{e_i e_j}}{\rho_{y_i e_j}} = \frac{1}{\kappa_i(x)} \qquad \frac{\rho_{e_i e_j}}{\rho_{e_i y_j}} = \frac{1}{\kappa_j(x)} \qquad \text{so } \frac{\rho_{y_i e_j}}{\rho_{e_i y_j}} = \frac{\kappa_i(x)}{\kappa_j(x)}.$$

TABLE I

Definition	Name
	Single predicted variable
$\rho_{x_i x_j}$	Correlation coefficient of zero order
ρ_{xy}	Correlation coefficient of zero order
$\rho_{xe} = 0$	None
$\rho_{xy} = \dfrac{\rho_{xy}}{\rho_{yx}}$	*Multiple augmented correlation coefficient
$\rho_{ye} = \kappa_y(x)$	Multiple alienation coefficient
$\rho_{yy} = \rho_y(x)$	Multiple correlation coefficient
$\rho_{ey} = 0$	None
	Two or more predicted variables
$\rho_{y_i y_j}$	Correlation coefficient of zero order
$\rho_{e_i e_j}$	Partial correlation coefficient
$\rho_{y_i y_j}$	*Predictions correlation coefficient
$\rho_{y_i e_j}$	Part correlation coefficient
$\rho_{y_i y_j}$	*Cross multiple correlation coefficient
$\rho_{e_i y_j}$	None

* Proposed name

Similarly

$$\text{(42)}\qquad \frac{\rho_{y_i y_j}}{\rho_{y_i y_j}} = \frac{\rho_i(x)}{\rho_j(x)}.$$

The geometric mean of similar coefficients yields such expressions as

$$\text{(43)}\qquad \begin{aligned} \sqrt{\rho_{y_i e_j}\, \rho_{e_i y_j}} &= \rho_{e_i e_j} \sqrt{\kappa_i(x)\, \kappa_j(x)} \\ \sqrt{\rho_{y_i y_j}\, \rho_{y_i y_j}} &= \rho_{y_i y_j} \sqrt{\rho_i(x)\, \rho_j(x)}. \end{aligned}$$

7. Determinantal formulas. The implicit normal equations (5) become when expanded

$$\text{(44)}\qquad \begin{aligned} \rho_{11}\beta_1 + \rho_{12}\beta_2 + \cdots + \rho_{1k}\beta_k &= \rho_{1y} \\ \rho_{21}\beta_1 + \rho_{22}\beta_2 + \cdots + \rho_{2k}\beta_k &= \rho_{2y} \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ \rho_{k1}\beta_1 + \rho_{k2}\beta_2 + \cdots + \rho_{kk}\beta_k &= \rho_{ky} \end{aligned}$$

while $\Sigma y\underline{y} = \Sigma\underline{y}^2 = \rho^2_{\underline{y}(x)}$ becomes

$$\rho_{y1}\beta_1 + \rho_{y2}\beta_2 + \cdots + \rho_{yk}\beta_k = \rho^2_{\underline{y}(x)}. \tag{45}$$

Let Δ be the determinant of the matrix of the solution of the k x's and y. Let Δ' be the corresponding determinant with ρ_{yy} replaced by $\rho^2_{\underline{y}(x)}$. Let Δ_{yy} be the determinant of the correlation-matrix of the k x's. Then $\rho^2_{\underline{y}(x)} = \Sigma\underline{y}^2 = \Sigma y\underline{y}$ can be expressed as a function of Δ and Δ_{yy} If (44) and (45) are to hold simultaneously, then $\Delta' = 0$. Expanding Δ' in terms of the bottom row, we get

$$\Delta' = 0 = \rho^2_{\underline{y}(x)}\Delta_{yy} + \text{"terms"}. \tag{46}$$

Similarly

$$\Delta = \rho_{yy}\Delta_{yy} + \text{"terms"} \tag{47}$$

where the "terms" of (46) and (47) are identical. It follows by subtraction that $\Delta = (1 - \rho^2_{\underline{y}(x)})\Delta_{yy}$ and hence that

$$\Sigma y\underline{y} = \Sigma\underline{y}^2 = \rho^2_{\underline{y}(x)} = 1 - \frac{\Delta}{\Delta_{yy}}. \tag{48}$$

Then

$$\Sigma e^2 = \Sigma e\underline{y} = \kappa^2_{\underline{y}(x)} = 1 - \Sigma\underline{y}^2 = 1 - \left(1 - \frac{\Delta}{\Delta_{yy}}\right) = \frac{\Delta}{\Delta_{yy}}. \tag{49}$$

Correlation formulas of section 3 then appear as

$$\rho_{x\underline{y}} = \frac{\rho_{xy}}{\sqrt{1 - \dfrac{\Delta}{\Delta_{yy}}}}, \tag{50}$$

$$\rho_{e\underline{y}} = \sqrt{\frac{\Delta}{\Delta_{yy}}}, \tag{51}$$

$$\rho_{y\underline{y}} = \sqrt{1 - \frac{\Delta}{\Delta_{yy}}}. \tag{52}$$

In a similar way the normal equations (25) become two sets of normal equations. The first set is like (44) with β_s replaced by $\beta_{y_i s}$ and ρ_{ys} replaced by $\rho_{y_i s}$. The second set is similar with i replaced by j. It is desired to find

$$\Sigma\underline{y}_i\underline{y}_j = \Sigma\underline{y}_i y_j = \rho_{y_j 1}\beta_1 + \rho_{y_j 2}\beta_2 + \cdots + \rho_{y_j k}\beta_k. \tag{53}$$

Now using (53) with (51) as applied to y_i and using the technique of the first part of this section, we get

$$\Delta_{y_i y_j} = \rho_{y_i y_j}\Delta_{y_i y_i\, y_j y_j} + \text{"terms"}, \tag{54}$$

$$0 = \Sigma y_i y_j \Delta_{y_i y_i\, y_j y_j} + \text{"terms"}, \tag{55}$$

where Δ is the determinant of the matrix of the correlations of the k x's, y_i and

y_j; $\Delta_{y_i y_j}$ is the determinant obtained by deleting the column involving correlations of y_i and the row involving correlations of y_i; $\Delta_{y_i y_i\, y_j y_j}$ is the determinant of the matrix of the k x's; and the "terms" in (54) and (55) are identical. It follows that

$$\Sigma \underline{y}_i \underline{y}_j = \rho_{ij} - \frac{\Delta_{y_i y_j}}{\Delta_{y_i y_i \cdot y_j y_j}} \tag{56}$$

and thence

$$\rho_{ij} - \Sigma \underline{y}_i \underline{y}_j = \frac{\Delta_{y_i y_j}}{\Delta_{y_i y_i \cdot y_j y_j}}. \tag{57}$$

The formulas of section (5) then appear in determinant form as follows

$$\rho_{e_i e_j} = \frac{\dfrac{\Delta_{ij}}{\Delta_{ii\,jj}}}{\sqrt{\left(\dfrac{\Delta_{ii}}{\Delta_{ii\,jj}}\right)\left(\dfrac{\Delta_{jj}}{\Delta_{jj\cdot ii}}\right)}} = \frac{\Delta_{ij}}{\sqrt{\Delta_{ii}\Delta_{jj}}} \tag{58}$$

as is well known.

$$\rho_{y_i y_j} = \frac{\rho_{ij} - \dfrac{\Delta_{ij}}{\Delta_{ii\,jj}}}{\sqrt{\left(1 - \dfrac{\Delta_{ii}}{\Delta_{ii\,jj}}\right)\left(1 - \dfrac{\Delta_{jj}}{\Delta_{jj\cdot ii}}\right)}}. \tag{59}$$

$$\rho_{y_i e_j} = \frac{\dfrac{\Delta_{ij}}{\Delta_{ii\,jj}}}{\sqrt{\dfrac{\Delta_{ij}}{\Delta_{ii\,jj}}}}. \tag{60}$$

$$\rho_{y_i y_j} = \frac{\rho_{ij} - \dfrac{\Delta_{ij}}{\Delta_{ii\,jj}}}{\sqrt{1 - \dfrac{\Delta_{ii}}{\Delta_{ii\,jj}}}}. \tag{61}$$

Formulas for $\rho_{e_i y_j}$ and $\rho_{y_i y_j}$ are similar to (60) and (61).

Modern methods of calculating determinants (2), (3), (4), (5) are advised if calculations are to be made from those formulas.

8. Matrix formulas. A matrix presentation is very useful in exhibiting the general features of this theory and in developing compact and easy methods of calculation with finite populations. The matrix presentation here is similar to that given by the author in a previous article [6].

Let the normal equations (24) be represented by the matrix equation.

$$E = Y - XB = Y - \underline{Y}. \tag{62}$$

Then the sets of normal equations become

$$X'E = 0 \quad \text{or} \quad X'(Y - XB) = 0$$

so that

$$X'XB = X'Y. \tag{63}$$

Now since $XB = \underset{\sim}{Y}$, (63) can be written as $X'\underset{\sim}{Y} = X'Y$ and it can be shown that

$$Y'Y = \underset{\sim}{Y}'Y = \underset{\sim}{Y}'\underset{\sim}{Y}. \tag{64}$$

But under the assumptions of section 2, $X'X$ is the matrix of the intercorrelations of the X's, $X'Y$ is the matrix of the intercorrelations of the x's and y's and $Y'Y$ is the matrix of the intercorrelations of the y's. Hence (63) can be written

$$R_{xx}B = R_{xy} \tag{65}$$

so that

$$B = R_{xx}^{-1}R_{xy}. \tag{66}$$

If Y is composed of a single variable, B is a single column matrix (vector) but if Y is composed of m variables, B is an m column matrix. It follows at once that

$$Y'Y = \underset{\sim}{Y}'\underset{\sim}{Y} = B'X'XB = B'R_{xy}B = R'_{xy}R_{xx}^{-1}R_{xx}R_{xx}^{-1}R_{xy} = R'_{xy}R_{xx}^{-1}R_{xy} \tag{67}$$

and that

$$\begin{aligned} E'E = (Y - XB)'E = Y'E = Y'(Y - XB) = Y'Y - Y'\underset{\sim}{Y} \\ = Y'Y - \underset{\sim}{Y}'\underset{\sim}{Y} = R_{yy} - R'_{xy}R_{xx}^{-1}R_{xy}. \end{aligned} \tag{68}$$

It thus appears that the matrix (67) has diagonal terms $\Sigma\underset{\sim}{y}^2 = \Sigma y\underset{\sim}{y}$ which are the squares of the multiple correlation coefficients, and that the non-diagonal terms are $\Sigma\underset{\sim}{y}_i\underset{\sim}{y}_j = \Sigma y_i\underset{\sim}{y}_j$. Similarly the matrix (68) has diagonal terms $\Sigma e^2 = \kappa^2_{y(x)}$ and non-diagonal terms $\Sigma e_i e_j = \Sigma e_i\underset{\sim}{y}_j$. It follows that all the correlation coefficients defined above may be calculated from the matrices R_{xx}, R_{xy}, R_{yy}, $\underset{\sim}{Y}'\underset{\sim}{Y}$, and $E'E$. The matrix (67) might be called the *multiple correlation matrix* and the matrix (68) the *multiple alienation matrix*.

Conventional results are expressed in terms of the correlation matrices R_{xx}, R_{xy}, and R_{yy}. All the correlation coefficients defined in this paper may be expressed in terms of these matrices and the multiple correlation and alienation matrices.

9. Calculational method of determining the multiple correlation and multiple alienation matrices. Various methods might be used in calculating the multiple correlation and alienation matrices from the correlation matrices. One method utilizes the square root method of solving simultaneous equations, which has

recently been presented in a number of places, [7] [8] together with a device which is similar to that used by Aitken [9] in eliminating the back solution This method solves the equation (65) by forming the auxiliary

$$S_{xx}B = S_{xx}R_{xx}^{-1}R_{xy} \tag{69}$$

where S_{xx} is a triangular matrix such that

$$R_{xx} - S'_{xx}S_{xx} = 0. \tag{70}$$

TABLE II

General		Illustration					
	R_{yy}					1.000	.495
						—	1.000
		1.000	.652	.554	.615	.313	.650
		—	1.000	.747	.693	.280	.803
R_{xx}	R_{xy}	—	—	1.000	.774	.182	.804
		—	—	—	1.000	.166	.812
		1.000	.652	.554	.615	.313	.650
			.758	.509	.385	.100	.500
S_{xx}	$S_{xx}R_{xx}^{-1}R_{xy}$			.659	.360	.064	.287
					.586	.072	.199
						.117	.221
	$Y'Y$					—	.794
						.883	.274
	$E'E$					—	.206

The right hand side of (69), when premultiplied by its transpose yields

$$(S_{xx}R_{xx}^{-1}R_{xy})'(S_{xx}R_{xx}^{-1}R_{xy}) = R'_{xy}R_{xx}^{-1}S'_{xx}S_{xx}R_{xx}^{-1}R_{xy} = R'_{xy}R_{xx}^{-1}R_{xy} = Y'Y. \tag{71}$$

Speaking less technically it is only necessary to multiply the columns of $S_{xx}R_{xx}^{-1}R_{xy}$ to get $Y'Y$.

A first illustration utilizes the correlations of the Carver anthropometric data [10] for 1000 University of Michigan freshmen. This group may be regarded as constituting a population, or it may be regarded as a random sample of a larger population. For present purposes we regard it as a population. Height (Y_1) and weight (Y_2) are estimated from shoulder girth (X_1) chest girth (X_2), waist girth (X_3), and right thigh girth (X_4) The calculation of $Y'Y$ and $E'E$ from the correlation matrices follow

As a second illustration I use the correlation between the parts of two forms of the Thorndike Intelligence Examination which Lorge has used in illustration canonical correlation technique [11, 69–74]. The X's are the scores on the three parts of Form A and the Y's are the scores on the three parts of Form B. In this case we designate the results by r's and k's (rather than ρ's and κ's) since the calculation is considered to be for a sample. The calculation of the sample multiple correlation and multiple alienation matrices is presented in Table III.

TABLE III

	Form A			Form B			
	x_1	x_2	x_3	y_1	y_2	y_3	
				1.0000	.8235	.7912	
				—	1.0000	.8315	R_{yy}
				—	—	1.0000	
	1.0000	.7830	.7852	.8986	.7841	.8217	
R_{xx}	—	1.0000	.8393	.7961	.8543	8254	R_{xy}
	—	—	1.0000	.7683	.8226	.8588	
	1.0000	.7830	.7852	.8986	.7841	.8217	
S_{xx}		.6220	.3609	.1487	.3864	2926	$S_{xx}R_{xx}^{-1}R_{xy}$
			.5032	.0180	.1341	.2146	
				.8299	.7645	.7858	
				—	.7821	.7861	$Y'Y$
				—	—	.8069	
				.1701	.0590	.0054	
				—	.2179	.0454	$E'E$
				—	—	.1991	

10. The numerical values of the coefficients. The diagonal entries of the multiple correlation matrix give the values of $\Sigma y_i^2 = \Sigma y_i \underline{y}_i = \rho_{y(x)}^2$ while the non-diagonal values are $\Sigma y_i y_j = \Sigma \underline{y}_i \underline{y}_j$. The diagonal entries of the multiple alienation matrix are $\Sigma e_i^2 = \Sigma e_i \underline{y}_i = \kappa_{y(x)}^2$ while the non-diagonal entries are $\Sigma e_i e_j = \Sigma e_i \underline{y}_j = \Sigma \underline{y}_i e_j$. We are then able to write out any of the correlations easily. Thus from Table II

$$\rho_{1(x)} = \sqrt{\Sigma \underline{y}_1^2} = \sqrt{.117} = .342,$$

$$\rho_{2(x)} = \sqrt{\Sigma \underline{y}_2^2} = \sqrt{.794} = .891,$$

$$\kappa_{1(x)} = \sqrt{\Sigma e_1^2} = \sqrt{.883} = .940,$$

$$\kappa_{2(x)} = \sqrt{\Sigma e_2^2} = \sqrt{206} = .454,$$

$$\rho_{12(x)} = \frac{\Sigma e_1 e_2}{\sqrt{(\Sigma e_1^2)(\Sigma e_2^2)}} = \frac{.274}{\sqrt{(.883)(.206)}} = .643,$$

$$\rho_{y_1 y_2} = \frac{\Sigma y_1 y_2}{\sqrt{(\Sigma y_1^2)(\Sigma y_2^2)}} = \frac{.221}{\sqrt{(.117)(.794)}} = .724,$$

$$\rho_{y_1 e_2} = \frac{\Sigma e_1 e_2}{\sqrt{\Sigma e_2^2}} = \frac{.274}{\sqrt{.206}} = .604,$$

$$\rho_{y_2 e_1} = \frac{\Sigma e_1 e_2}{\sqrt{\Sigma e_1^2}} = \frac{.274}{\sqrt{.883}} = .291,$$

$$\rho_{y_1 y_2} = \frac{\Sigma y_1 y_2}{\sqrt{\Sigma y_2^2}} = \frac{.221}{\sqrt{.794}} = .248,$$

$$\rho_{y_1 y_2} = \frac{\Sigma y_1 y_2}{\sqrt{\Sigma y_1^2}} = \frac{.221}{\sqrt{.117}} = .646.$$

TABLE IVa

General			Illustration		
	$r_{y_1 y_2}$	$r_{y_1 y_3}$		.9489	.9603
$\rho_{1(x)}$	$r_{y_1 y_2} \mid r_{y_1 y_2}$	$r_{y_1 y_3} \mid r_{y_1 y_3}$	.9110	.8392 | .8644	.8626 | .8747
Σy_1^2	$\Sigma y_1 y_2$	$\Sigma y_1 y_3$	.8299	.7645	.7858
		$r_{y_2 y_3}$			.9917
	$r_{2(x)}$	$r_{y_2 y_3} \mid r_{y_2 y_3}$		.8844	.8889 | .8751
	Σy_2^2	$\Sigma y_2 y_3$		.7821	.7861
		$r_{3(x)}$			.8983
		Σy_3^2			.8069

TABLE IVb

General			Illustration		
	$r_{e_1 e_2}$	$r_{e_1 e_3}$		.3066	.0298
$k_{1(x)}$	$r_{e_1 y_2} \mid r_{y_1 e_2}$	$r_{e_1 y_3} \mid r_{y_1 e_3}$	.4124	.1431 | .1264	.0131 | .0123
Σe_1^2	$\Sigma e_1 e_2$	$\Sigma e_1 e_3$	.1701	.0590	.0054
		$r_{e_2 e_3}$			.2214
	$k_{2(x)}$	$r_{e_2 y_3} \mid r_{y_2 e_3}$		.4668	.0973 | .1033
	Σe_2^2	$\Sigma e_2 e_3$		.2179	.0454
		$k_{3(x)}$			.4394
		Σe_3^2			.1931

It is possible to utilize a scheme of successive division if all these correlations are desired when there are more than two predicted variables. By divisions we compute in turn $\rho_{1(x)}$, $\rho_{y_i y_j}$, $\rho_{y_i y_j}$ and $\rho_{y_i y_j}$ from the multiple correlation matrix and $\kappa_{i(x)}$ $\rho_{e_i y_j}$, $\rho_{y_i e_j}$, $\rho_{e_i e_j}$ from the multiple alienation matrix for each i, j. The computational scheme is illustrated in Table IV where the correlations used are the sample correlations of Table III. The calculations from the multiple correlation matrix are presented in Table IVa and those from the multiple alienation matrix in Table IVb.

In Table IVa the multiple correlation matrix is first entered on the third of each three lines The square root of each diagonal term is then extracted to give the multiple correlation coefficients The value of $r_{1(x)}$ is then locked in the machine as a divisor and it is divided, in turn, into $\Sigma y_1 y_2$, $\Sigma y_1 y_3$ to get $r_{y_1 y_2}$ and $r_{y_1 y_3}$. Then $r_{2(x)}$ is used as a divisor by division into $r_{y_1 y_2}$ to get $r_{y_1 y_2}$, into $\Sigma y_1 y_2$ to get $r_{y_1 y_2}$, and into $\Sigma y_2 y_3$ to get $r_{y_2 y_3}$. Finally $r_{3(x)}$ is divided into $r_{y_1 y_3}$ to get $r_{y_1 y_3}$, into $\Sigma y_1 y_3$ to get $r_{y_1 y_3}$, into $r_{y_2 y_3}$ to get $r_{y_2 y_3}$ and into $\Sigma y_2 y_3$ to get $r_{y_2 y_3}$. A check on these divisions can be made, if desired, by dividing $r_{y_1 y_2}$ by $r_{1(x)}$ to get $r_{y_1 y_2}$, $r_{y_1 y_3}$ by $r_{1(x)}$ to get $r_{y_1 y_3}$, and $r_{y_2 y_3}$ by $r_{2(x)}$ to get $r_{y_2 y_3}$.

Table IVb is treated in a similar manner.

This technique is immediately applicable to the case of many predicted variables.

REFERENCES

[1] M. Ezekiel, *Methods of Correlation Analysis*, Second Edition, Wiley, 1942.

[2] A. C. Aitken, "On the evaluation of determinants, the formation of their adjugates, etc.," *Edinb. Math. Soc. Proc.*, Series 2, Vol 3(1933), pp 207-219.

[3] P. S. Dwyer, "The evaluation of determinants," *Psychometrika*, Vol. 6(1941), pp 191-204.

[4] A. C. Aitken, *Determinants and Matrices*, Second Edition, Oliver and Boyd, Edinburgh, 1942.

[5] F. V Waugh and P. S. Dwyer, "Compact computation of the inverse of a matrix," *Annals of Math. Stat*, Vol. 16(1945), pp. 359-371

[6] P. S. Dwyer, "A matrix presentation of least squares and correlation theory, etc.," *Annals of Math. Stat.*, Vol 15(1944), pp. 82-89.

[7] P. S. Dwyer, "The square root method and its use in correlation and regression," *Am. Stat. Assn Jour*, Vol. 40(1945), pp. 493-503.

[8] D. B. Duncan and J. F. Kenney, *On the Solution of Normal Equations and Related Topics*, Edwards Brothers, Ann Arbor, 1946.

[9] A. C. Aitken, "The evaluation of a certain triple product matrix," *Roy Soc. of Edinb. Proc.*, Vol. 57(1937), pp. 172-181.

[10] H. C Carver, *Anthropometric Data*, Edwards Brothers, Ann Arbor, 1941

[11] Irving Lorge, "The computation of Hotelling canonical correlations," *Proceedings of Educational Research Forum*, Endicott, N. Y, Aug. 26-31, 1940, pp. 68-74.

INVERSION FORMULAS IN NORMAL VARIABLE MAPPING

By John Riordan

Bell Telephone Laboratories, New York

1. Summary. The two inversion formulas considered here arise from study of G. A. Campbell's work on the Poisson summation, which is described more fully in the introduction and in the main consists of finding a function or mapping of a variable connected with the summation in terms of a normal (Gaussian) variable g. More generally, this last is a process often called "normalization of the variable" and associated with the names of E. A. Cornish and R. A. Fisher. The mapping is two-way and the main inversion formula determines co-efficients for one way from those for the other, both sets of coefficients being descriptive of their mappings More precisely if x is a given variable, g a Gaussian variable, y a parameter of the mapping, and the two mappings are

$$x = g + \sum_1^\infty G_n(g)\, y^n/n!,$$

$$g = x + \sum_1^\infty X_n(x) y^n/n!,$$

the formula expresses $G_n(x)$ in terms of $X_i(x)$, $i \leq n$, and vice versa.

The second formula is more particularly related to the Poisson summation and relates coefficients $p_n \equiv p_n(g)$ and $q_n \equiv q_n(g)$ in the pair of equations

$$a = c \sum_0^\infty q_n c^{-\frac{1}{2}n}/n!$$

$$c = a \sum_0^\infty p_n a^{-\frac{1}{2}n}/n!$$

Both formulas, which are necessarily elaborate, are given concise expression by the use of the multi-variable polynomials of E. T. Bell.

2. Introduction. In 1923, in a paper little known in statistical circles, G. A. Campbell [2] gave as the basis for his extensive tabulation of the Poisson summation an asymptotic series expressing the average a in terms of a normal variable g, corresponding to the probability of at least c occurrences, and c itself. That is to say, he associated with the Poisson summation

$$P(a, c) = \sum_c^\infty e^{-a} a^x/x!$$

a normal variable g, defined by

$$P(a, c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g} e^{-x^2/2}\, dx$$

and inverted the summation (which, as is well known, is equivalent to the incomplete Gamma function ratio) to give a series for a in terms of g and c. The series, which is carried to 11 terms, starts as follows:

$$a \sim c\left[1 + gc^{-1/2} + \frac{g^2 - 1}{3}\, c^{-1} + \frac{g^3 - 7g}{36}\, c^{-3/2} + \cdots\right]$$

If $x = (a - c)\, c^{-1/2}$ is introduced, this becomes

$$x \sim g + \frac{g^2 - 1}{3}\, c^{-1/2} + \frac{g^3 - 7g}{36}\, c^{-1} + \cdots$$

and x is seen to be, like g, a standardized variable of mean 0, variance 1.

It seems to have gone unnoticed that this result includes the χ^2 distribution through the transformation: $2a = X^2$, $2c = n$ and it has been rediscovered by A. M. Peiser [7] (4 terms) and by Goldberg and Levine [4] (6 terms).

It is possible also to express c in terms of a and g, and a formula of this kind with fewer terms which appears in a footnote in Campbell's paper is as follows:

$$c \sim a\left[1 - ga^{-1/2} + \frac{g^2 + 2}{6}\, a^{-1} + \frac{g^3 + 2g}{72}\, a^{-3/2} + \cdots\right]$$

Finally there is a third possibility of expressing g in terms of the remaining variables, preferably x and c; though unnoticed by Campbell this has since been brought to prominence by Cornish and Fisher [3], Hotelling and Frankel [5] and Kendall [6].

The idea behind the first expansion appears most clearly in the second form and is that for c large the variable x behaves nearly like g. The third possibility reverses this expansion and gives a function of x and c which behaves like g; hence if this function is first evaluated, reference to the normal integral table gives an immediate evaluation of the probabilities in question. Put in another way, the expansion widens the scope of the normal integral table and for this reason has been called "normalization" of the variable (but this term seems preempted by its use in another sense for orthogonal functions, and has been replaced in the title by normal variable mapping).

From the point of view of statistical theory, the three expressions are different versions of one relationship, which suggests that there should be general rules for transforming a series of one type into that of another. The two inversion formulas given below supply these rules in what appears to be as compact a form as the problem allows. It will be noted that the proofs given suppose convergent series, a case which leads to clarity and brevity and is interesting in itself. Applied to Campbell's series, they give the known results so far as the latter go, but of course for other asymptotic series they need independent verifications.

3. First Inversion Formula. This relates coefficients in series like Campbell's first and its reverse as in Cornish and Fisher. More precisely

If $G_1(g)$, $G_2(g)$ $\cdots$ are assigned polynomials and if

$$x = g + \sum_{n=1}^{\infty} G_n(g)y^n/n!, \tag{1}$$

defines x in terms of g and a parameter y, then

$$g = x + \sum_{1}^{\infty} X_n(x)y^n/n!, \tag{2}$$

where

$$-X_n(x) = Y_n(aG_1(x), aG_2(x), \cdots, aG_n(x)), \tag{3}$$

TABLE 1

Bell Polynomials Y_n $(fg_1, fg_2 \cdots fg_n)$

$$\begin{aligned}
Y_1 &= f_1g_1 \\
Y_2 &= f_1g_2 + f_2g_1^2 \\
Y_3 &= f_1g_3 + f_2(3g_2g_1) + f_3g_1^3 \\
Y_4 &= f_1g_4 + f_2(4g_3g_1 + 3g_2^2) + f_3(6g_2g_1^2) + f_4g_1^4 \\
Y_5 &= f_1g_5 + f_2(5g_4g_1 + 10g_3g_2) + f_3(10g_3g_1^2 + 15g_2^2g_1) \\
&\quad + f_4(10g_2g_1^3) + f_5g_1^5 \\
Y_6 &= f_1g_6 + f_2(6g_5g_1 + 15g_4g_2 + 10g_3^2) \\
&\quad + f_3(15g_4g_1^2 + 60g_3g_2g_1 + 15g_2^3) \\
&\quad + f_4(20g_3g_1^3 + 45g_2^2g_1^2) + f_5(15g_2g_1^4) + f_6g_1^6 \\
Y_7 &= f_1g_7 + f_2(7g_6g_1 + 21g_5g_2 + 35g_4g_3) \\
&\quad + f_3(21g_5g_1^2 + 105g_4g_2g_1 + 70g_3^2g_1 + 105g_3g_2^2) \\
&\quad + f_4(35g_4g_1^3 + 210g_3g_2g_1^2 + 105g_2^3g_1) \\
&\quad + f_5(35g_3g_1^4 + 105g_2^2g_1^3) + f_6(21g_2g_1^5) + f_7g_1^7 \\
Y_8 &= f_1g_8 + f_2(8g_7g_1 + 28g_6g_2 + 56g_5g_3 + 35g_4^2) \\
&\quad + f_3(28g_6g_1^2 + 168g_5g_2g_1 + 280g_4g_3g_1 + 210g_4g_2^2 + 280g_3^2g_2) \\
&\quad + f_4(56g_5g_1^3 + 420g_4g_2g_1^2 + 280g_3^2g_1^2 + 840g_3g_2^2g_1 + 105g_2^4) \\
&\quad + f_5(70g_4g_1^4 + 560g_3g_2g_1^3 + 420g_2^3g_1^2) \\
&\quad + f_6(56g_3g_1^5 + 210g_2^2g_1^4) + f_7(28g_2g_1^6) + f_8g_1^8
\end{aligned}$$

Y_n being the multivariable polynomial of E. T. Bell [1], *in the variables $G_1(x)$ to $G_n(x)$ and the symbolic variable a which is such that*

$$a^i \equiv a_i = (-D)^{i-1}, \qquad D = d/dx,$$

with differentiations on all products of $G_1(x)$ to $G_n(x)$ associated with it in the polynomial.

Note the symmetry of x and g, which allows the transformation to go either way, the inverse of (3) being

$$-G_n(g) = Y_n(aX_1(g), aX_2(g) \cdots, aX_n(g)) \tag{4}$$

Table I gives explicit expressions for polynomials Y_1 to Y_8. It will be noted that the number of terms in Y_n is the number of partitions of n and that f_s, the

variable replacing a_i in the table, is associated with terms corresponding to partitions with i parts; that is to say, if $Y_{n,i}$ designates such terms

$$Y_n = \sum_1^n f_i Y_{n,i}$$

The verification or extension of the table may be accomplished by the formulas and relations given by Bell (l.c.) or more directly by those modifications of Bell given by myself in [8].

The first few instances of (3), dropping the common variable x for brevity, may be read off from Table I (with appropriate changes of notation and interpretation of a_i) as follows:

$$-X_1 = G_1$$
$$-X_2 = G_2 - D(G_1^2)$$
$$-X_3 = G_3 - 3D(G_2G_1) + D^2(G_1^3)$$
$$-X_4 = G_4 - 4D(G_3G_1) - 3D(G_2^2) + 6D^2(G_2G_1^2) - D^3(G_1^4)$$

Applied to Campbell's first formula in its second form with $y = c^{-1/2}$ and

$$G_1(x) = (x^2 - 1)/3, \qquad G_3(x) = (-6x^4 - 14x^2 + 32)/270,$$
$$G_2(x) = (x^3 - 7x)/18, \qquad G_4(x) = (9x^5 + 256x^3 - 433x)/1680,$$

these show e.g.

$$-X_2 = \frac{x^3 - 7x}{18} - \frac{2(x^2 - 1)}{3} \cdot \frac{2x}{3} = \frac{-7x^3 + x}{18},$$

and similarly for the others, resulting in

$$X_1 = -(x^2 - 1)/3$$
$$X_2 = (7x^3 - x)/18$$
$$X_3 = -(219x^4 - 14x^2 - 13)/270$$
$$X_4 = (3993x^5 - 152x^3 + 119x)/1680$$

These determine a calculation formula for the Poisson summation, which is a refinement of the normal approximation. That is to say

$$P(a, c) = \Phi(g) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g} e^{-t^2/2}\, dt$$

with

$$g = x - \frac{x^2 - 1}{3\sqrt{c}} + \frac{7x^3 - x}{36c} - \frac{219x^4 - 14x^2 - 13}{1620c\sqrt{c}} + \frac{3993x^5 - 152x^3 + 119x}{40320c^2} - \cdots$$

and $x = (a - c)/\sqrt{c}$.

For the t-variate, the formula is applied in the reverse direction since Hotelling and Frankel supply the first four values of X_n, that is, in present notation, the series

$$g \sim x - \frac{x^3 + x}{4} y + \frac{13x^5 + 8x^3 + 3x}{48} \frac{y^2}{2} - \frac{35x^7 + 19x^5 + x^3 - 15x}{64} \frac{y^3}{6} + \frac{6271x^9 + 3224x^7 - 102x^5 - 1680x^3 - 945x}{3840} \frac{y^4}{24} + \cdots$$

The reversed series (obtained by (4)) is

$$x \sim g + \frac{g^3 + g}{4} y + \frac{5g^5 + 16g^3 + 3g}{48} \frac{y^2}{2} + \frac{3g^7 + 19g^5 + 17g^3 - 15g}{64} \frac{y^3}{6} + \frac{79g^9 + 776g^7 + 1482g^5 - 1920g^3 - 945g}{3840} \frac{y^4}{24} + \cdots$$

The first three terms are checked by Goldberg and Levine (l.c.).

Another application worth noting is to the formulas of Cornish and Fisher which give $G_r(g)$ and $X_r(x)$ in terms of the relative cumulants of the distribution; to save space these are omitted.

The derivation of the formula may be indicated most easily by Lagrange's formula for the expansion of one function in powers of another in the following form[1]:

Let C be a contour in the complex z plane enclosing the point $z = x$, and let $f(z)$ and $\phi(z)$ be analytic on and inside C. Let y be such that $| y\phi(z) | < | z - x |$ when z is on C, and g be that root of the equation:

$$g = x + y\phi(g) \tag{5}$$

which lies inside C. Then

$$f(g) = \frac{1}{2\pi i} \int_C f(z) \frac{d}{dz} \{\log [z - x - y\phi(z)]\}\, dz = f(x) + \sum_1^\infty X_n^*(x) y^n/n! \tag{6}$$

where

$$X_n^*(x) = \frac{d^{n-1}}{dx^{n-1}} [f'(x)(\phi(x))^n] \tag{7}$$

The contour integral in (6) appears, slightly disguised, as a problem in Whittaker and Watson [*Modern Analysis*, Cambridge, 1920, p. 149]. The evaluation (7) is given for completeness, though no use is made of it in this section, the derivation proceeding directly from (6).

First notice that by (1) and (5)

$$-y\phi(g) = \sum_1^\infty G_n(g) y^n/n!,$$

[1] The author owes the suggestion for this to S O Rice, who also simplified the derivation of the second inversion formula given later.

so that the logarithm in (6) may be written

$$\log (z - x + \sum_1^\infty G_n(z)y^n/n!),$$

or

$$\log (z - x) + \log \left[1 + \sum_1^\infty G_n(z)(z - x)^{-1}y^n/n!\right],$$

or

$$\log (z - x) + \log \exp by, \tag{8}$$

with b a symbolic variable such that

$$b^0 \equiv b_0 = 1$$

$$b^n \equiv b_n = G_n(z)(z - x)^{-1}.$$

Now if

$$\log (\exp by) = B_1y + B_2y^2/2! + \cdots, \tag{9}$$

$$= \exp By,$$

B being another symbolic variable, $B_0 = 0$, $B^n \equiv B_n$, it follows from equation (5) of [8] that

$$B_n = [D_y^n \log (\exp by)]_{y=0}, \qquad D_y = d/dy,$$

$$= Y_n(\beta b_1, \beta b_2, \cdots \beta b_n)$$

$$= \sum_1^n \beta_i Y_{n,i}(b_1, b_2, \cdots b_n),$$

with $\beta_i = (-)^{i-1}(i - 1)!$ and $Y_{n,i}$ the part of polynomial Y_n having i parts, as defined above. Moreover, each factor b_k of terms in $Y_{n,i}$ contributes $G_k(z)(z - x)^{-1}$ so that

$$B_n = \sum_1^n \beta_i(z - x)^{-i} Y_{n,i}(G_1(z), G_2(z) \cdots G_n(z)) \tag{11}$$

Then, by (5)

$$f(g) = \frac{1}{2\pi i}\int_C f(z)\left(\frac{1}{z - x} + \frac{d}{dz}\exp By\right)dz$$

$$= f(x) - \frac{1}{2\pi i}\int_C f'(z)\exp By\, dz$$

$$= f(x) - \frac{1}{2\pi i}\int_C f'(z)\sum_1^\infty \frac{y^n}{n!}\sum_1^n \frac{\beta_i Y_{n,i}(G_1 \cdots G_n)}{(z - x)^i}\, dz$$

$$= f(x) - \sum_1^\infty \frac{y^n}{n!} \sum_{i=1}^n \int_C \frac{(-)^{i-1}(i-1)!}{2\pi i(z-x)^i} Y_{n,i}(G_1(z) \cdots G_n(z)) f'(z)\, dz$$

$$= f(x) - \sum_1^\infty \frac{y^n}{n!} \sum_1^n (-D)^{i-1}[f'(x) Y_{n,i}(G_1(x) \cdots G_n(x))]$$

with $D = d/dx$. The evaluation in the last line is by the Cauchy formula for derivates; the second line is derived by an integration by parts.

Equation (4) follows from this and the substitution $f(g) = g$.

4. Second Inversion Formula. This gives the interrelations of coefficients of series like the two Campbell series mentioned in the introduction. It runs as follows:

If $q_1(g)$, $q_2(g)$ $\cdots$ are given polynomials and if

$$a = c \sum_0^\infty \frac{q_n(g) c^{-\frac{1}{2}n}}{n!} \tag{12}$$

defines a in terms of g and a parameter c; then

$$c = a \sum_0^\infty \frac{p_n(g) a^{-\frac{1}{2}n}}{n!} \tag{13}$$

where

$$-p_n(g) = Y_n(\alpha q_1(g),\ \alpha q_2(g),\ \cdots,\ \alpha q_n(g)) \tag{14}$$

with $\alpha^1 \equiv \alpha_1 = 1$; $\alpha^i \equiv \alpha_i = (n-4)(n-6)\cdots(n-2i)2^{-i+1}$

Equation (14) is formally similar to (3) and by symmetry as before, $q_n(g)$ is readily expressible as a Y_n polynomial in $p_1(g)$ to $p_n(g)$.

The first five instances of (14), dropping the argument for brevity, are

$$-p_1 = q_1$$

$$-p_2 = q_2 - q_1^2$$

$$-p_3 = q_3 - \tfrac{3}{2} q_2 q_1 + \tfrac{3}{4} q_1^3$$

$$-p_4 = q_4$$

$$-p_5 = q_5 + \tfrac{5}{2}(q_4 q_1 + 2q_3 q_2) - \tfrac{5}{4}(2q_3 q_1^2 + 3q_2^2 q_1) + \tfrac{15}{4} q_2 q_1^3 - \tfrac{15}{16} q_1^5$$

Applied to Campbell's first series where

$$q_1(g) = g \qquad q_3(g) = (g^3 - 7g)/6$$

$$q_2(g) = \tfrac{2}{3}(g^2 - 1) \qquad q_4(g) = (-12g^4 - 28g^2 + 64)/135$$

$$q_5(g) = (36g^5 + 1024g^3 - 1732g)/1296$$

these show that

$$p_1(g) = -g \qquad p_3(g) = (g^3 + 2g)/12$$
$$p_2(g) = (g^2 + 2)/3 \qquad p_4(g) = (12g^4 + 28g^2 - 64)/135$$
$$p_5(g) = (207g^5 + 2596g^3 - 6148g)/1296$$

The proof of (14) is as follows. First, for brevity introduce symbolic variables p and q with the usual interpretation $p^n \equiv p_n(g)$, $q^n \equiv q_n(g)$ so that (12) and (13) read

$$a = c \exp q\, c^{-\frac{1}{2}}$$
$$c = a \exp p\, a^{-\frac{1}{2}}$$

Now write $a = 1/x^2$, $c = 1/y^2$ changing these to

$$x = y\ (\exp qy)^{-\frac{1}{2}}$$
$$y = x\ (\exp px)^{-\frac{1}{2}}$$

and note that

$$x^2y^{-2} = (\exp qy)^{-1} = \exp px \tag{15}$$

which shows that p_n is the coefficient of $x^n/n!$ in the expansion in powers of x of $(\exp qy)^{-1}$. Lagrange's formula gives at once ($D = d/dy$):

$$f(y) = \sum_1^\infty \frac{x^n}{n!} D^{n-1}[f'(g)(\exp qy)^{\frac{1}{2}n}]_{y=0} \tag{16}$$

so that

$$\begin{aligned}(\exp qy)^{-1} &= \sum_1^\infty \frac{x^n}{n!} D^{n-1}[-(\exp qy)^{\frac{1}{2}(n-4)} D(\exp qy)]_{y=0} \\ &= \sum_1^\infty \frac{x^n}{n!} D^{n-1}\left[-\frac{2}{n-2} D(\exp qy)^{\frac{1}{2}(n-2)}\right]_{y=0} \\ &= \sum_1^\infty \frac{x^n}{n!}\left(\frac{-2}{n-2}\right)[D^n(\exp qy)^{\frac{1}{2}(n-2)}]_{y=0}\end{aligned}$$

or

$$\begin{aligned}-p_n &= \frac{2}{n-2}[D^n(\exp qy)^{\frac{1}{2}(n-2)}]_{y=0} \\ &= Y_n(\alpha q_1, \alpha q_2, \cdots, \alpha q_n)\end{aligned} \tag{17}$$

with α, as in (14), by equation (5) of [8].

BIBLIOGRAPHY

[1] Bell, E T., "Exponential polynomials," *Annals of Math.* Vol. *35*, (1934), pp. 258–277.
[2] Campbell, G. A., "Probability curves showing Poisson's exponential summation,"

Bell System Tech Jl. Vol *2*, (1923), pp. 95–113; Collected Papers, New York, 1937, pp 224–242.

[3] Cornish, E. A and Fisher, R A., "Moments and cumulants in the specification of distributions," *Revue de l'Institut Intern de Stat* Vol. *4*, (1937), pp. 1–14.

[4] Goldberg, H and Levine, H., "Approximate formulas for the percentage points and normalization of t and χ^2," *Annals of Math Stat* Vol *17*, (1946), pp. 216–225

[5] Hotelling, H. and Frankel, L R , "The transformation of statistics to simplify their distribution," *Annals of Math Stat.* Vol *9*, (1938), pp 87–96.

[6] Kendall, M G., *The Advanced Theory of Statistics* I, London. 1943.

[7] Peiser, A. M , "Asymptotic formulas for significance levels of certain distributions," *Annals of Math Stat* Vol *14*, (1943), pp 56–62

[8] Riordan, John, "Derivatives of Composite Functions," *Bull. Am. Math. Soc.* Vol *52*, (1946), pp 664–667.

ON THE DETERMINATION OF OPTIMUM PROBABILITIES IN SAMPLING

By Morris H. Hansen and William N. Hurwitz

Bureau of the Census

1. Summary. In a previous paper [2] it was shown that it is sometimes profitable to select sampling units with probability proportionate to size of the unit. This note indicates a method of determining the probabilities of selection which minimize the variance of the sample estimate at a fixed cost. Some approximations that have practical applications are given.

2. Introduction. Neyman has shown that it is possible to reduce the sampling variance of an estimate by dividing a population into sub-populations (called strata) and varying the proportions of units included in the sample from stratum to stratum [1]. His treatment presumed that the units within any stratum would be drawn with equal probability. In many practical sampling problems, the use of constant probabilities is neither necessary nor desirable. Not only is it possible to obtain unbiased or consistent estimates with varying probabilities of selection of the sampling units, but also it is possible to reduce the variance of sample estimates by appropriate use of this device.

It has been shown [2] that in a subsampling system, the selection of primary units with probabilities proportionate to the number of elements included in the primary unit may bring about marked reductions in sampling variances over sampling with equal probabilities. In this note, we shall indicate a method of determining the optimum probabilities under certain conditions, and also some approximations to the optima that have practical applications.

By optimum probabilities, we mean the set of probabilities of selection that will minimize the variance for a fixed cost of obtaining sample results, or alternatively that will minimize the cost for a fixed sampling error.

3. Optimum probability with a subsampling system. Consider, for example, the simple subsampling system where primary units are first drawn for inclusion in the sample and then a sample of elements is drawn from the selected primary units. We shall suppose, for simplicity of notation, that the sampling is done without stratification. The conclusions indicated below will be similar if stratified sampling is used, and they will hold even if only one unit is drawn from each stratum. Suppose that a population contains M primary units, and that the sampling of primary units is to be done with replacement. Sampling with replacement is assumed in order to simplify the mathematics. We wish to estimate the ratio

$$\frac{X}{Y} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N_i} X_{ij}}{\sum_{i=1}^{M} \sum_{j=1}^{N_i} Y_{ij}}$$

where X_{ij} and Y_{ij} are the values of two characteristics of the jth element within the ith primary unit, and N_i is the number of elements in the ith primary unit. A consistent estimate of X/Y is given by

$$r = \frac{\sum_{i=1}^{m} \frac{N_i}{P_i} \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^{m} \frac{N_i}{P_i} \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}} \tag{1}$$

where

P_i = The probability of selecting the ith primary unit on a single draw

n_i = The total number of elements included in the sample from the ith unit if it is drawn. If a particular unit happens to be included in the sample more than once the subsampling will be independently carried through each time it is drawn.

m = The total number of primary units included in the sample.

It will be assumed that a self-weighting sample is to be used, i.e., that although the probabilities of selecting primary units will vary, the subsampling rate within the ith selected primary unit, $\frac{n_i}{N_i}$, will be such that $P_i \frac{n_i}{N_i} = k$. Note that, with this condition, k is the probability that an element will be included in the sample by making a single draw of a primary unit, and by carrying out the specified subsampling within the selected primary unit. It follows that mkN is the expected total number of elements included in a sample of primary units, where

$$N = \sum_{i=1}^{M} N_i.$$

The method can be extended to cover situations where other conditions are imposed.

We shall express the variance of r in terms of P_i, m, and k, and also express the cost in terms of these same quantities. The optimum values of P_i, m, and k will then be determined

The variance of the sample estimate. To terms of order $1/m$ of the Taylor expansion of a ratio, the sampling variance of the estimate (1) is approximately

$$\sigma_r^2 \doteq \frac{\sum_{i=1}^{M} \frac{N_i^2}{P_i} \Delta_i^2 + \sum_{i=1}^{M} \frac{N_i^2}{P_i} \frac{N_i - n_i}{N_i n_i} \sigma_i^2}{mY^2} \tag{2}$$

where

$$\Delta_i^2 = \bar{Y}_i^2 \left(\frac{\bar{X}_i}{\bar{Y}_i} - \frac{X}{Y} \right)^2, \qquad \bar{Y}_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}, \qquad \bar{X}_i = \frac{\sum_{j=1}^{N_i} X_{ij}}{N_i},$$

$$\sigma_i^2 = \sigma_{ix}^2 + \frac{\bar{X}^2}{\bar{Y}^2}\sigma_{iy}^2 - 2\frac{\bar{X}}{\bar{Y}}\sigma_{ixy},$$

$$\sigma_{ix}^2 = \frac{\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2}{N_i - 1},$$

$$\sigma_{iy}^2 = \frac{\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2}{N_i - 1},$$

$$\sigma_{ixy} = \frac{\sum_{j=1}^{N_i} X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i)}{N_i - 1}.$$

The cost function. Now suppose that the total cost of the sampling procedure involves a fixed cost attached to each primary unit included in the sample, a cost of listing the elements within each selected primary unit (this listing may be necessary in order to draw a subsample), and a cost of obtaining information from each of the elements selected for inclusion in the sample. Under these circumstances the total expected cost of the survey will be:

$$C = C_1 m + C_2 m \sum_{i=1}^{M} P_i N_i + C_3 mkN \tag{3}$$

where

C_1 = The fixed cost per primary unit,
C_2 = The cost of listing one element in a selected primary unit and other costs that vary with the number of elements to be listed,
C_3 = The cost of obtaining the required information from one element in the sample,
$\sum_{i=1}^{M} P_i N_i$ = Expected number of elements in the sample per primary unit in the sample,
mk = The over-all sampling ratio, and
$N = \sum_{i=1}^{M} N_i$ = The total number of elements in the population.

It will be noted that although the values of P_i and m may be fixed in advance, the number of elements to be listed, $\sum_{i=1}^{m} N_i$, remains a chance variable. It is for this reason that we consider the expected cost rather than the actual cost.

The optimum values of P_i, m, and k. The values of P_i, m, and k which minimize the variance (2) subject to the conditions that:

$$C \text{ is fixed}, \qquad \frac{n_i}{N_i} P_i = k, \qquad \sum_{i=1}^{M} P_i = 1,$$

are given by

$$P_i = \frac{\sqrt{\frac{N_i^2 \delta_i}{C_1 + C_2 N_i}}}{\sum_{i=1}^{M} \sqrt{\frac{N_i^2 \delta_i}{C_1 + C_2 N_i}}}, \tag{4}$$

$$k = \frac{\sqrt{\frac{\sum_{i=1}^{M} N_i \sigma_i^2}{N}}}{\sum_{i=1}^{M} \sqrt{\frac{N_i^2 \delta_i}{C_1 + C_2 N_i}}\, C_3}, \tag{5}$$

$$m = \frac{C}{C_1 + C_2 \sum_{i=1}^{M} P_i N_i + C_3 kN}, \tag{6}$$

where

$$\delta_i = \Delta_i^2 - \frac{\sigma_i^2}{N_i}.$$

Ordinarily δ_i will be positive although it will often be found to be negative for some i. For a great many populations, such negative values can be avoided by classifying the primary units into size groups or other significant groups and then requiring that the probability of selection be P_α for every primary unit in the α-th group.

In actual practice, however, in advance of designing a sample one does not have the data to compute the optima and uses methods of approximating the optimum probabilities. Methods of approximating the optimum probabilities are given below.

4. Some rules for approximating the optimum probabilities. In another paper [2] considerations were presented from which it follows that δ_i tends to decrease with increasing size of unit, but seldom as fast as the size of unit increases. The rate of decrease is often small relative to the increase in N_i, and empirical data for a number of problems indicate that even the assumption of δ_i being fairly constant with increasing size of unit may not lead one far astray from the optimum probabilities. Under this assumption ($\delta_i = \delta$ for all i) the probabilities depend only on N_i, C_1, and C_2, and lead to the following results:

(a) When $C_1 > 0$ and $C_2 = 0$, probability proportionate to size will be the optimum.

(b) When $C_1 = 0$ and $C_2 > 0$, probability proportionate to the square root of the size will be the optimum.

If we go to the other extreme (extreme not in terms of mathematically possible values but in terms of most practical populations), and assume that δ_i decreases at the same rate that N_i increases, the results would be:

(a) When $C_1 > 0$ and $C_2 = 0$, probability proportionate to the square root of the size will be the optimum.

(b) When $C_1 = 0$ and $C_2 > 0$, equal probability will be the optimum.

The minimum is broad in the neighborhood of the optimum and the results for either of these extremes and the values in between often will give results reasonably close to the minimum. This leads to the following useful approximations:

(a) When $C_2 \Sigma P_i N_i$, the expected cost per primary unit of listing and related operations, is small in relation to C_1, the fixed cost per primary unit, the optimum probabilities will be between probability proportionate to size and probability proportionate to the square root of size, and either of these will be reasonably close to the optimum.

(b) When C_1 is small compared to $C_2 \Sigma P_i N_i$, the optimum probability will be between equal probability and probability proportionate to the square root of size, and either of these will be reasonably close to the optimum.

(c) When both C_1 and $C_2 \Sigma P_i N_i$ are of significant size, i.e., when the costs vary substantially both with the number of primary units in the sample and the size of the units, then probability proportionate to the square root of the size will be a reasonably good approximation to the optimum.

(d) When units of small size are used and all of the subunits in the selected primary units are included in the sample (that is, there is no subsampling) equal probability is close to the optimum. It should be noted that this rule does not follow directly from the above analysis based on subsampling, but from a separate analysis in which no subsampling is involved.

For whatever system of probabilities is used, and with the cost function given by (3), the optimum value of k is given by:

$$k = \sqrt{\frac{\sum_{i=1}^{M} N_i \sigma_i^2 \left(C_1 + C_2 \sum_{j=1}^{M} P_j N_j \right)}{C_3 N \left(\sum_{i=1}^{M} \frac{N_i^2 \Delta_i^2}{P_i} - \sum_{i=1}^{M} N_i \sigma_i^2 \right)}},$$

which can be approximated, in application, from prior experience or preliminary studies. The corresponding optimum value for m is obtained by substitution in the cost function

The above results should not be accepted, of course, as the optima for every cost function or every sampling system. Either past experimental data may be available or pilot tests made to determine the cost function and the appropriate approximations that should be used in various practical situations.

An illustration An illustration may be of interest. A characteristic published for city blocks in the 1940 Census of Housing is the number of dwelling units that are in need of major repairs or that lack a private bath. Suppose we

were sampling to estimate the proportion of the dwelling units having this characteristic for the Bronx in New York City, at the time of the 1940 Census. Let us assume that once we selected a system of probabilities we used the optimum numbers of blocks and the optimum sampling ratios appropriate to these probabilities, that is, the optimum values of k and m. For each of several cost functions the following Table 1 shows the sampling variances of each system, relative to the variance of sampling with equal probability. It also shows values of $C_2\Sigma P_i N_i$ for comparison with C_1.

TABLE 1

Unit costs			Average cost per primary unit of listing and related operations ($C_2\Sigma P_i N_i$)			Variances relative to equal probability		
C_1	C_2	C_3	Equal probability	Probability proportionate to square root of size	Probability proportionate to size	Equal probability	Probability proportionate to square root of size	Probability proportionate to size
5	.10	1	13.49	21.15	27.63	100	92	104
5	.05	1	6.75	10.58	13.82	100	88	97
5	.02	1	2.70	4.23	5 53	100	83	87
5	0	1	0	0	0	100	75	73
2	.10	1	13.49	21.15	27.63	100	96	111
2	.05	1	6.75	10.58	13.82	100	93	106
2	.02	1	2.70	4.23	5 53	100	90	97
2	0	1	0	0	0	100	79	77
1	.10	1	13.49	21.15	27.63	100	97	114
1	.05	1	6.75	10.58	13 82	100	96	110
1	.02	1	2.70	4 23	5 53	100	93	103
1	0	1	0	0	0	100	82	81
0	.10	1	13.49	21.15	27.63	100	99	117
0	.05	1	6.75	10.58	13 82	100	99	115
0	.02	1	2.70	4 23	5.53	100	99	113

Some of the costs given in the table do not have unreasonable relationships in terms of the situations encountered in practice in various types of jobs. The comparisons are not affected by the absolute magnitudes of the costs but only by their relative magnitudes The results are consistent with the rough rules of thumb given above. It is worth noting that in each of the above instances probability proportionate to the square root of the size yields a comparatively low variance

5. Sampling with or without replacement. In this paper the sampling with varying probabilities was assumed to be carried out with replacement which ordinarily would not be advisable in practice. When sampling is done without replacement the optimum probabilities and their approximations will be about the same as for sampling with replacement in at least those instances where the proportion of the population in the sample is small. Further investigation is needed for large sampling rates.

6. Conclusion. In summary, it is not essential and may not be desirable to give each element in the population (or stratum) the same chance of being drawn in order to avoid bias or to have a consistent estimate. Estimate (1) is a consistent estimate no matter what probabilities of selection are assigned to these units. The use of variable probabilities of selection is another device to be added to those already in the literature, such as stratification and efficient methods of estimation, which make it possible to achieve the objectives of a sample survey at reduced costs. Reference [2] gives another illustration of reductions in sampling variance achieved through the use of varying probabilities in accordance with the rules suggested above for approximating the optimum probabilities.

REFERENCES

[1] JERZY NEYMAN, "On the two different aspects of the representative method of purposive selection," *Roy. Stat. Soc Jour.*, New Series, Vol. 97 (1934), pp. 558-606.

[2] MORRIS H. HANSEN AND WILLIAM N. HURWITZ, "On the theory of sampling from finite populations," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 333-362

A SOLUTION TO THE PROBLEM OF OPTIMUM CLASSIFICATION

By P. G. Hoel and R. P. Peterson

University of California, Los Angeles

1. Summary. By means of a general theorem, the space of the variables of classification is separated into population regions such that the probability of a correct classification is maximized. The theorem holds for any number of populations and variables but requires a knowledge of population parameters and probabilities. A second theorem yields a large sample criterion for determining an optimum set of estimates for the unknown parameters. The two theorems combine to yield a large sample solution to the problem of how best to discriminate between two or more populations.

2. Introduction. There are essentially two basic problems in discriminant analysis. The first problem is to test whether the populations differ, since it would be futile to attempt a classification if the populations did not differ. The second problem is to find an efficient method for classifying individuals into their proper populations. In this paper, an optimum asymptotic solution of the second problem will be presented.

3. Parameters known. Let $f_i = f_i(x_1, \cdots, x_k)$, $(i = 1, \cdots, r)$ denote the probability density function of population i in the region under consideration. Let $p_i > 0$, $(i = 1, \cdots, r)$, denote the probability that population i will be sampled if a single individual is selected at random from that region, and let R denote the k dimensional Euclidean variable space Then the desired theorem is the following:

Theorem 1. *If M_i denotes the region in R where $p_if_i \geq p_jf_j$, $(j = 1, \cdots, r)$, and where $p_if_i > 0$, then the set of regions M_i, $(i = 1, \cdots, r)$, in which any overlap is assigned to the M_i with the smallest index, will maximize the probability of a correct classification.*

For the purpose of proving this theorem, consider any other set of non-overlapping regions, M_i'. Since the addition to any of the regions M_i of a part of R throughout which all the functions f_i vanish will not affect the probability of a correct classification, there is no loss of generality in assuming that the set of regions M_i' contains the same portion of R as the set of regions M_i does. The relationship between the two sets may be expressed by means of the formulas

$$M_i = \sum_{j=1}^{r} M_{ij} \tag{1}$$

and

$$M_j' = \sum_{i=1}^{r} M_{ij}, \tag{2}$$

where M_{ij} denotes that part of M_i which is contained in M_j'.

Since a sample point that falls in the region M_i will be judged to have come from population i, the probability of the correct classification of a single random sample by means of the set M_i is given by

$$Q = p_1 \int_{M_1} f_1 dE + \cdots + p_r \int_{M_r} f_r dE, \tag{3}$$

where $dE = dx_1 dx_2 \cdots dx_r$. If Q' denotes the probability of the correct classification by means of the set M_i',

$$Q' = p_1 \int_{M_1'} f_1 \, dE + \cdots + p_r \int_{M_r'} f_r \, dE.$$

In the notation of (1) and (2), these probabilities become

$$Q = p_1 \int_{\sum_j M_{1j}} f_1 \, dE + \cdots + p_r \int_{\sum_j M_{rj}} f_r \, dE$$

and

$$Q' = p_1 \int_{\sum_i M_{i1}} f_1 \, dE + \cdots + p_r \int_{\sum_i M_{ir}} f_r \, dE.$$

Now consider the difference $Q - Q'$. It can be expressed in the form

$$\begin{aligned} Q - Q' &= \sum_{i=1}^{r} \sum_{j=1}^{r} \left[p_i \int_{M_{ij}} f_i \, dE - p_j \int_{M_{ij}} f_j \, dE \right] \\ &= \sum_{i=1}^{r} \sum_{j=1}^{r} \int_{M_{ij}} [p_i f_i - p_j f_j] \, dE. \end{aligned}$$

Since M_{ij} is contained in M_i and $p_i f_i \geq p_j f_j$, $(j = 1, \cdots, r)$, holds throughout M_i, it follows that each of these integrals is non-negative, consequently $Q \geq Q'$, which proves the theorem.

This theorem yields a solution to the classification problem only when the f_i are completely specified and the p_i are known.

It will be observed that this theorem is similar to a generalization of a fundamental lemma in the Neyman-Pearson theory of testing hypotheses [1], and to a result by Welch [2].

If the basic weight function in Wald's [3] formulation of the multiple decision problem assumes only the values 0 and 1, corresponding to whether or not a correct classification is made, it will be found that the set of regions M_i will minimize the expected value of the loss in that formulation.

4. Parameters unknown. Since the p_i, as well as the parameters in the f_i, are assumed to be unknown, Q will be a function of such parameters. Let $\theta_1, \cdots, \theta_s$ denote all such parameters, including the p_i. Now let a random sample of size n be taken from the region under consideration and let $\bar{\theta}_1, \cdots, \bar{\theta}_s$ denote a set of

estimates of the parameters based on this sample. Since the total sample will constitute a sample of size n_1 from f_1, n_2 from f_2, etc, where $n = n_1 + \cdots + n_r$, the θ's for f_i will be estimated by means of a sample of size n_i rather than of size n. In the following arguments, it will not be necessary to distinguish between θ's which are estimated by different size samples because the arguments will be based on the order of terms with respect to the size sample and $n_i \sim np_i$ with probability one. Or, more simply, choose all n_i equal.

Let $\bar{M}_i$ correspond to M_i when the parameters are replaced by their sample estimates and let $\bar{Q}$ denote the probability of a correct classification when using the regions $\bar{M}_i$ in place of the regions M_i. Then, from (3),

$$Q - \bar{Q} = \sum_{i=1}^{r} p_i \left[\int_{M_i} f_i \, dE - \int_{\bar{M}_i} f_i \, dE \right].$$

Let $H = Q - \bar{Q}$ Since the estimates, $\bar{\theta}_i$, are random variables, H will be a random variable which is a function of the estimation functions, $\bar{\theta}_i$, as well as of the parameters, θ_i. The desired criterion for determining optimum estimates is then given by the following theorem:

THEOREM 2. *If* $E(\bar{\theta}_i - \theta_i)^4 = O(n^{-g})$, $g > 0$, *and if in some neighborhood of the point* $\bar{\theta}_i = \theta_i$, $(i = 1, \cdots, s)$ *the function* H *is continuous and possesses continuous derivatives of the first, second, and third order with respect to the* $\bar{\theta}_i$, *then*

$$E(H) = \frac{1}{2} \sum_{i=1}^{s} \sum_{j=1}^{s} H_{ij} E(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) + O(n^{-3/4g}),$$

where H_{ij} *denotes the partial derivative of* H *with respect to* $\bar{\theta}_j$ *and* $\bar{\theta}_i$ *at the point* $(\theta_1, \cdots, \theta_s)$

The proof is similar to the type of proof used by Cramer [4] to obtain an expression for the variance of a function of central moments.

By means of Tchebycheff's inequality [4], page 182, it follows that

$$P[(\bar{\theta}_i - \theta_i)^4 \geq \epsilon^4] \leq \frac{E(\bar{\theta}_i - \theta_i)^4}{\epsilon^4}.$$

From the theorem assumptions, there exists a constant A such that

$$P[\bar{\theta}_i - \theta_i)^4 \geq \epsilon^4] < \frac{An^{-g}}{\epsilon^4}.$$

This is equivalent to

$$P[|\bar{\theta}_i - \theta_i| \geq \epsilon] < \frac{An^{-g}}{\epsilon^4}.$$

If E_1 denotes the set of points in sample space where $|\bar{\theta}_i - \theta_i| < \epsilon$, $(i = 1, \cdots, s)$, and E_2 denotes the complementary set, this inequality implies that

$$P[E_2] < \frac{sAn^{-g}}{\epsilon^4} \tag{4}$$

The expected value of H may be written in the form

$$E(H) = \int_{E_1} H\, dP + \int_{E_2} H\, dP. \tag{5}$$

Consider the order of the second integral. From (4) and the fact that H is the difference of two probabilities, it follows that

$$\left| \int_{E_2} H\, dP \right| \leq \int_{E_2} dP = P[E_2] < \frac{sA\, n^{-g}}{\epsilon^4}.$$

Consequently (5) becomes

$$E(H) = \int_{E_1} H\, dP + O(n^{-g}). \tag{6}$$

Now consider the first integral. From the theorem assumptions, if ϵ is chosen sufficiently small, it follows that for any point in the set E_1, the function H can be expanded in the form

$$H = H(\theta) + \sum_1^s (\bar{\theta}_i - \theta_i) H_i(\theta) + \frac{1}{2} \sum_1^s \sum_1^s (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) H_{ij}(\theta) + R,$$

where θ denotes the point $(\theta_1, \cdots, \theta_s)$, where

$$R = \frac{1}{6} \sum_1^s \sum_1^s \sum_1^s (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)(\bar{\theta}_k - \theta_k) H_{ijk}(\theta'),$$

and where θ' is some point in E_1. Since $\bar{Q}$ reduces to Q when $\bar{\theta} = \theta$, $H(\theta) = 0$. Furthermore, since Q denotes the maximum probability of a correct classification, $H \geq 0$ for all $\bar{\theta}$; hence $H_i(\theta) = 0$ and $H_{ii}(\theta) \geq 0$ for all i. Thus, for any point in the set E_1,

$$H = \frac{1}{2} \sum_1^s \sum_1^s (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) H_{ij}(\theta) + R.$$

If this expression is substituted in (6), $E(H)$ will become

$$E(H) = \frac{1}{2} \sum_1^s \sum_1^s H_{ij}(\theta) \int_{E_1} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)\, dP + \int_{E_1} R\, dP + O(n^{-g}). \tag{7}$$

Consider, first, the order of the remainder term. From the continuity assumption on H_{ijk}, it follows that H_{ijk} is bounded in E_1, say $|H_{ijk}(\theta')| < B$; hence

$$\left| \int_{E_1} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)(\bar{\theta}_k - \theta_k) H_{ijk}(\theta')\, dP \right| < B \int_{E_1} |(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)(\bar{\theta}_k - \theta_k)|\, dP.$$

By Schwarz's inequality,

$$\int_{E_1} |(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)(\bar{\theta}_k - \theta_k)|\, dP \leq \left[\int_{E_1} (\bar{\theta}_i - \theta_i)^2 (\bar{\theta}_j - \theta_j)^2\, dP \int_{E_1} (\bar{\theta}_k - \theta_k)^2\, dP \right]^{\frac{1}{2}}.$$

Similarly,

$$\int_{E_1} (\bar{\theta}_i - \theta_i)^2(\bar{\theta}_j - \theta_j)^2\, dP \leq \left[\int_{E_1} (\bar{\theta}_i - \theta_i)^4\, dP \int_{E_1} (\bar{\theta}_j - \theta_j)^4\, dP\right]^{\frac{1}{2}},$$

$$\int_{E_1} (\bar{\theta}_k - \theta_k)^2\, dP \leq \left[\int_{E_1} (\bar{\theta}_k - \theta_k)^4\, dP \int_{E_1} dP\right]^{\frac{1}{2}} \leq \left[\int_{E_1} (\bar{\theta}_k - \theta_k)^4\, dP\right]^{\frac{1}{2}}.$$

Since

$$\int_{E_1} (\bar{\theta}_i - \theta_i)^4\, dP \leq \int_{E_1+E_2} (\bar{\theta}_i - \theta_i)^4\, dP = O(n^{-g}),$$

the preceding inequalities combine to give

$$\left|\int_{E_1} R\, dP\right| = O(n^{-3/4g}). \tag{8}$$

Now consider the first integral in (7). It may be written in the form

$$\int_{E_1} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)\, dP = E(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) - \int_{E_2} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)\, dP. \tag{9}$$

By Schwarz's inequality,

$$\left|\int_{E_2} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)\, dP\right| \leq \left[\int_{E_2} (\bar{\theta}_i - \theta_i)^2\, dP \int_{E_2} (\bar{\theta}_j - \theta_j)^2\, dP\right]^{\frac{1}{2}}.$$

Similarly,

$$\int_{E_2} (\bar{\theta}_i - \theta_i)^2\, dP \leq \left[\int_{E_2} (\bar{\theta}_i - \theta_i)^4\, dP \cdot P[E_2]\right]^{\frac{1}{2}}.$$

If these inequalities are combined and inequality (4) is employed, (9) will reduce to

$$\int_{E_1} (\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j)\, dP = E(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) + O(n^{-g}). \tag{10}$$

Finally, if (8) and (10) are employed in (7), it will reduce to the result stated in the theorem.

The order of the leading term in $E(H)$ depends upon the nature of the estimating functions, $\bar{\theta}_i$. In order to insure that this term will be the dominating term, and thus rule out pathological situations, only that class of estimating functions (estimators) will be considered for which this term will be of lower order than that of the remainder term. If the estimators are means or central moments, for example, then $g = 2$. For such estimators the order of the remainder term is $O(n^{-\frac{1}{2}})$, whereas the order of the leading term is not higher than $O(n^{-1})$.

A set of estimators will be called an optimum set if it maximizes the expected value of the probability of a correct classification, or, what is equivalent, if it minimizes $E(H)$. Since only large samples are being considered here, it is neces-

sary to define optimum in an asymptotic sense. Consider sets of estimators for which $E(H)$ is of order $O(n^{-q})$. For this class of estimators, a set will be called asymptotically optimum if it minimizes

$$\lim_{n\to\infty} n^q E(H)$$

Among asymptotically optimum sets of various orders, the set corresponding to the highest order would naturally be considered as the best asymptotic set. Now from Theorem 2, it readily follows that a set of estimators which minimizes

$$\sum_{1}^{s}\sum_{1}^{s} H_{ij} E(\bar{\theta}_i - \theta_i)(\bar{\theta}_j - \theta_j) \tag{11}$$

will be an asymptotically optimum set.

5. Maximum likelihood estimates. If the estimates $\bar{\theta}_i$ are unbiased and uncorrelated, (11) will reduce to

$$\sum_{1}^{s} H_{ii}\sigma_i^2 \tag{12}$$

where $\sigma_i^2 = E(\bar{\theta}_i - \theta_i)^2$ is a function of n as well as of the parameters. Since, from the discussion preceding (7), $H_{ii} \geq 0$, it follows that (12) will be a minimum when the σ_i^2 assume their minimum values. Now it is known [4], page 504, that under mild restrictions maximum likelihood estimates possess minimum asymptotic variances; hence for estimators of the type being considered which also satisfy the conditions in [4], the maximum likelihood estimates of the θ_i will yield an asymptotically optimum set of estimates for the classification problem.

REFERENCES

[1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. Phil. Trans.*, Vol. 231 (1933), pp. 289–337.

[2] B. L. Welch, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218–220.

[3] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses," *Annals of Math Stat.*, Vol. 10 (1939), pp. 299–304.

[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 352–356.

NOTES

This section is devoted to brief research and expository articles on methology and other short items.

A GENERALIZATION OF WALD'S FUNDAMENTAL IDENTITY

By Gunnar Blom

University of Stockholm

1. Summary. The fundamental identity is generalized to the case of independent random variables with non-identical distributions The conditions for the validity of the differentiation of the identity are discussed. The results given in [1], [2], and [3] are obtained as special cases.

2. A property of cumulative sums. Let $z_1, z_2, \cdots$ be an infinite sequence of independent random variables, $F_1(z), F_2(z), \cdots$ their distribution functions (d.f.) and $\varphi_1(t), \varphi_2(t), \cdots$ their moment-generating functions so that $\varphi_\nu(t) = E(e^{tz_\nu})$. a_N and b_N are given constants ($a_N > b_N$, $N = 1, 2, \cdots$). n is defined as the smallest integer N for which $Z_N = z_1 + \cdots + z_N$ is $\geqq a_N$ or $\leqq b_N$.

We first give two lemmas.

Lemma 1. *If two positive quantities δ and ϵ can be found such that one at least of the following conditions a) and b) are satisfied*

$$a)\ P(z_\nu > \delta) > \epsilon \text{ for all } \nu \text{ and } \lim_{N\to\infty} \sup a_N < \infty$$

$$b)\ P(z_\nu < -\delta) > \epsilon \text{ for all } \nu \text{ and } \lim_{N\to\infty} \inf b_N > -\infty,$$

then for any $k \geqq 0$

$$\lim_{N\to\infty} N^k P(n > N) = 0. \tag{1}$$

An inspection of the proof of (4) in [4] shows that this formula holds when the conditions of the lemma are satisfied. The lemma follows.

Lemma 1 can be generalized as follows.

Lemma 2. *If two positive quantities δ and ϵ and a sequence $c_1, c_2, \cdots$ can be found such that one at least of the following conditions a) and b) are satisfied*

$$a)\ P(z_\nu + c_\nu > \delta) > \epsilon \text{ for all } \nu, \quad \lim_{N\to\infty} \sup a_N < \infty, \quad \lim_{N\to\infty} \sup \sum_1^N c_\nu < \infty,$$

$$b)\ P(z_\nu + c_\nu < -\delta) > \epsilon \text{ for all } \nu,$$

$$\lim_{N\to\infty} \inf b_N > -\infty, \quad \lim_{N\to\infty} \inf \sum_1^N c_\nu > -\infty,$$

then (1) *is true.*

PROOF: In case a) we put $z_\nu' = z_\nu + c_\nu$, $Z_N' = \Sigma z_\nu'$ and $a_N' = a_N + \Sigma_1^N c_\nu$. The inequality $Z_N \geqq a_N$ then becomes $Z_N' \geqq a_N'$. As $P(z_\nu' > \delta) > \epsilon$ and $\limsup_{N\to\infty} a_N' < \infty$, Lemma 1 can be applied to the sequence $z_1', z_2', \cdots$, and thus (1) is true. When conditions b) are satisfied, the proof is analogous.

3. The generalized fundamental identity. In this section we shall consider sequences of random variables of the type defined in Lemma 2. We shall prove two theorems the first of which is valid for complex values of t and the second only for real values of t.

THEOREM 1. *Assuming that*

1°. *one at least of conditions a) and b) of Lemma 2 is satisfied;*

2°. $b \leqq b_N < a_N \leqq a$, *where a and b are finite,*

3°. *for some complex (or real) value of t, $\varphi_\nu(t)$ exists for all ν and is $\neq 0$ and*

$$\liminf_{N\to\infty} |\varphi_1(t)\cdots\varphi_N(t)| > 0,$$

then

$$E[e^{tZ_n}(\varphi_1(t)\cdots\varphi_n(t))^{-1}] = 1. \tag{2}$$

PROOF. Let W_m denote the set of all sequences $z_1 \cdots z_N$ in the N-dimensional Euclidean space Ω_N for which $n = m$ ($m \leqq N$), W_m' the projection of W_m on Ω_m and $W_{n>N}$ all sequences for which $n > N$. We have identically

$$\left[\sum_{m=1}^{N}\int_{W_m} + \int_{W_{n>N}}\right] e^{tZ_N}\, dF_1 \cdots dF_N = \int_{\Omega_N} e^{tZ_N}\, dF_1 \cdots dF_N = \varphi_1(t)\cdots\varphi_N(t).$$

Dividing by the right member and cancelling common factors we obtain

$$\sum_{m=1}^{N} (\varphi_1\cdots\varphi_m)^{-1}\int_{W_m'} e^{tZ_m}\, dF_1\cdots dF_m + (\varphi_1\cdots\varphi_N)^{-1}\int_{W_{n>N}} e^{tZ_N}\, dF_1\cdots dF_N = 1. \tag{3}$$

When $N \to \infty$ the first sum tends to the left member of (2). We thus have to investigate the last term in (3) which we denote by R_N. We can write

$$R_N = (\varphi_1\cdots\varphi_N)^{-1}\int_{W_{n>N}} e^{tZ_N}\, dF_1\cdots dF_N = (\varphi_1\cdots\varphi_N)^{-1} P(n > N) E_{n>N} e^{tZ_N}. \tag{4}$$

It follows from Lemma 2 that $P(n > N) \to 0$. As $b < Z_N < a$ by 2° we conclude that $R_N \to 0$. This proves the theorem.

THEOREM 2. *If, for some real value of t, $\varphi_\nu(t)$ exists for all ν and if quantities c_ν, $\epsilon > 0$ and $\delta > 0$ can be found such that at least one of the following conditions a) and b) are satisfied for all ν*

a) $\limsup_{N\to\infty} a_N < \infty$, $\limsup_{N\to\infty} \sum_1^N c_\nu < \infty$ *and*

$$A_\nu(t,\delta) = \frac{1}{\varphi_\nu(t)}\int_{\delta-c_\nu}^{\infty} e^{tz}\, dF_\nu(z) > \epsilon, \qquad (\nu = 1, 2, \cdots), \tag{5a}$$

b) $\liminf_{N\to\infty} b_N > -\infty$, $\liminf_{N\to\infty} \sum_1^N c_\nu > -\infty$ *and*

$$B_\nu(t, \delta) = \frac{1}{\varphi_\nu(t)} \int_{-\infty}^{-\delta - c_\nu} e^{tz}\, dF_\nu(z) > \epsilon, \qquad (\nu = 1, 2, \cdots), \tag{5b}$$

then (2) *holds.*

The conditions of the theorem become more attractive if the theorem is limited to the somewhat less general cases mentioned in the Corollary below. The above formulation has been chosen mainly because of an important application to identical variables in Sec. 6.

PROOF. The theorem is proved if we can show that R_N in (4) tends to zero when $N \to \infty$. For that purpose we use the transformation (cf [5] and [3])

$$G_\nu(z; t) = \frac{1}{\varphi_\nu(t)} \int_{-\infty}^{z} e^{tz}\, dF_\nu(z), \qquad (\nu = 1, 2, \cdots). \tag{6}$$

$G_\nu(z; t)$ is obviously a d f. for every real t (for which $\varphi_\nu(t)$ exists). When (5a) holds,

$$P[z_\nu + c_\nu > \delta \mid G_\nu(z; t)] = A(t, \delta).$$

Here the expression in the left member denotes the probability that $z_\nu + c_\nu > \delta$, when G_ν is the d.f. of z_ν.

Consequently, when conditions a) are fulfilled, a sequence of random variables with the d.f:s $G_1(z; t)$, $G_2(z; t)$, $\cdots$ or, with one notation, $G(t)$ satisfies the conditions a) of Lemma 2. It follows that

$$\lim_{N\to\infty} P(n > N \mid G(t)) = 0.$$

Introducing $G_\nu(z; t)$ in R_N we find

$$R_N = \int_{W_n > N} dG_1 \cdots dG_N = P(n > N \mid G(t)).$$

Consequently $R_N \to 0$. When conditions b) are fulfilled, the proof is analogous.

COROLLARY TO THEOREM 2. *If* 1° $\varphi_\nu(t)e^{tc_\nu} \leqq H(t) < \infty$, 2° t *is positive and conditions a) of Lemma 2 hold or* t *is negative and conditions b) of Lemma 2 hold, then the generalized fundamental identity is true.*

For, in the first case

$$A_\nu(t, \delta) \geqq \frac{e^{t(\delta - c_\nu)}}{\varphi_\nu(t)} \int_{\delta - c_\nu}^{\infty} dF_\nu \geqq \frac{\epsilon e^{t\delta}}{H(t)} = \epsilon_1(t)$$

so that (5a) is satisfied, and similarly when t is negative.

The following special case deserves particular attention as it covers most cases occurring in practice and the conditions become very simple: *If a sequence of random variables satisfies conditions a) and b) of Lemma 1 simultaneously, a sufficient condition for the validity of* (2) *for some given real value of* t *is that the sequence* $\varphi_\nu(t)$ *is bounded.*

4. Application to Poisson variables. As an application of (2) we consider a sequence of Poisson variables with the parameters λm_ν, where λ is a positive quantity and m_ν are positive integers. From the well-known formula

$$\varphi_\nu(t) = e^{\lambda m_\nu (e^t - 1)}$$

we easily conclude that the conditions of Theorem 1 are valid if $R(e^t) \geqq 1$. (With $\delta < 1$ in (5a) we find that (2) holds even for negative t.) If, in particular, we choose t so that $e^t = 1 + \frac{2\pi i k}{\lambda} = c_k$, we have the simple formula

$$E(c_k^{z_n}) = 1, \qquad (k = 1, 2, \cdots).$$

5. Differentiation of the generalized fundamental identity. In this section t is assumed to be real. We denote the kth derivative of $\varphi_\nu(t)$ by $\varphi_\nu^{(k)}(t)$. We shall prove the following theorem which corresponds to Theorems 1 and 2.

THEOREM 3. *If for all t in a closed interval I the conditions stated in Theorems 1 or 2 are satisfied and if, in addition, the functions* $\left|\frac{\varphi_\nu^{(k)}(t)}{\varphi_\nu(t)}\right|$ *are uniformly bounded with respect to both ν and t (in I) for $k = 1, 2, \cdots r$, then the generalized fundamental identity may be differentiated r times with respect to t for any t in the interior of I*

We use a method of proof which is similar to that used in [2]. We first show that the sum in (3) may be differentiated r times under the integral signs and secondly that the rth derivative of R_N tends to zero uniformly in t when $N \to \infty$.

The rth derivative of the general term of the series in (3) consists of a finite number of terms of the form

$$J_m(t) = (\varphi_1 \cdots \varphi_m)^{-1} H_\mu \int_{W_m'} Z_m^\lambda e^{tz_m}\, dF_1 \cdots dF_m \qquad (\mu \leqq \lambda; \mu, \lambda = 1, 2, \cdots r),$$

and the rth derivative of R_N in (4) consists of a finite number (which does not depend on N) of similar expressions with N substituted for m and $W_{n>N}$ for W_m'. H_μ is a sum of m^μ and N^μ terms respectively which is symmetric in ν. The terms are functions of $\frac{\varphi_\nu^{(k)}(t)}{\varphi_\nu(t)}$ $(k \leqq \lambda; \nu = 1, 2, \cdots m)$ and are thus majorated by the same constant C.

Further, we can always find a positive quantity t_0 such that for all t in I

$$|Z_m^\lambda e^{tz_m}| \leqq e^{t_0|z_m|} \leqq (e^{t_0 z_m} + e^{-t_0 z_m}).$$

Hence

$$|J_m(t)| \leqq (\varphi_1 \cdots \varphi_m)^{-1} C m^\mu \int_{W_m'} (e^{t_0 z_m} + e^{-t_0 z_m})\, dF_1 \cdots dF_m \tag{7}$$

The rest of the proof is divided into two parts corresponding to the conditions of Theorem 1 and those of Theorem 2.

When the conditions of Theorem 2 are fulfilled we make the transformation (6) in (7) with $t = t_0$ and $t = -t_0$. Then

$$| J_m(t) | \leqq Cm^\mu[P(n = m \mid G(t_0)) + P(n = m \mid G(-t_0))] \leqq 2Cm^\mu < \infty.$$

This justifies the differentiation of the series in (3).

Substituting N for m and $n > N$ for $n = m$ in the above expression we further have

$$| J_N(t) | \leqq CN^\mu[P(n > N \mid G(t_0)) + P(n > N \mid G(-t_0))],$$

and conclude from Lemma 2 with $k = \mu$ in (1) that $J_N(t)$ tends to zero uniformly in t. It follows that the rth derivative of R_N also tends to zero uniformly in t.

In the second part of the proof we assume the conditions of Theorem 1 to be satisfied. We then write (7) in the following form

$$| J_m(t) | \leqq C(\varphi_1 \cdots \varphi_m)^{-1} m^\mu P(n = m) E_{n=m}(e^{t_0 Z_m} + e^{-t_0 Z_m}), \tag{8}$$

where $E_{n=m}$ signifies the conditional expectation when it is known that $n = m$. From the definition of n it follows that, when $n = m$, we have $b_{m-1} < Z_{m-1} < a_{m-1}$ and $Z_m \geqq a_m$ or $\leqq b_m$. Hence

$$E_{n=m}(e^{t_0 Z_m}) \leqq E_{n=m}(e^{t_0 Z_m} \mid Z_m \geqq a_m) = E_{n=m}[e^{t_0(Z_{m-1}+z_m)} \mid Z_{m-1} + z_m \geqq a_m]$$
$$\leqq e^{t_0 a_{m-1}} E[e^{t_0 z_m} \mid z_m > a_m - b_{m-1}] < \infty.$$

The second exponential can be treated in a similar way. Thus $J_m(t)$ is majorated by a finite expression

Finally, we substitute N for m and $n > N$ for $n = m$ in (8). I being a closed interval it follows from condition 3° in Theorem 1 that we can find a constant C such that

$$| J_N(t) | \leqq CN^\mu P(n > N) E_{n>N}(e^{t_0 Z_N} + e^{-t_0 Z_N}).$$

From the definition of n and condition 2° in Theorem 1 we have $b < Z_N < a$. An application of Lemma 2 then shows that $J_N(t)$ tends to zero uniformly in t. This proves the theorem.

COROLLARY TO THEOREM 3. *When the conditions stated in Corollary of Theorem 2 are fulfilled for all t in the closed interval I, Theorem 3 is true.*

This is obvious.

6. The fundamental identity for identically distributed variables. In the special case of identically distributed variables for which $P(z = 0) < 1$ and $0 < \varphi(t) < \infty$ we infer from Theorem 1 that the fundamental identity

$$E[e^{tZ_n}(\varphi(t))^{-n}] = 1 \tag{9}$$

holds if t is complex and $| \varphi(t) | \geqq 1$. This is the case discussed in [1].

Further, when $P(z = 0) < 1$, the integrals $\int_\alpha^\infty e^{tz} dF$ and $\int_{-\infty}^\beta e^{tz} dF$ cannot both

be zero for every $\alpha > 0$ and $\beta < 0$, and thus we infer from Theorem 2 that the fundamental identity holds for all real t (if the limits a_N and b_N are chosen in accordance with the conditions of this theorem). This proposition is somewhat more general than that proved in [3] by a similar method.

It also follows from the last remark and Theorem 3 that, when $P(z = 0) < 1$, (9) can be differentiated any number of times for any real t. This proposition contains the results in [2] and [3] as special cases.

7. A generalization. We finally remark that the assumption made in Theorem 3 that the expressions containing derivatives of $\varphi_r(t)$ are uniformly bounded is unnecessarily restrictive. For example, it seems possible to prove that the first derivative of (2) may be obtained by differentiation under the expectation sign if the series (cf. Corollary 1 to Theorem 7.4. in [6])

$$\sum_{m=1}^{\infty} P(n = m) \sum_{r=1}^{m} \frac{\varphi_r'(t)}{\varphi_r(t)}$$

is uniformly convergent with respect to t.

REFERENCES

[1] A. Wald, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), p. 285.

[2] A. Wald, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 493–497.

[3] G. E. Albert, "A note on the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 593-596 and Vol. 19 (1948), pp. 426–427.

[4] C. Stein, "A note on cumulative sums," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.

[5] H. Cramér, Sur un nouveau theorème-limite de la théorie des probabilités, *Actualités, scientifiques et industrielles*, no. 736, Hermann et Cie., 1938, p. 5.

[6] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 228–229.

SPREAD OF MINIMA OF LARGE SAMPLES

By Brockway McMillan

Bell Telephone Laboratories, Murray Hill, N. J.

1. Theorems. Let x have the continuous cumulative distribution function $F(x)$. Let $(x_1, \cdots, x_N)$ be a sample of N independent values of x and $y = \inf (x_1, \cdots, x_N)$. Then y is a random variable with the cumulative distribution function

$$G_N(y) = 1 - (1 - F(y))^N. \tag{1}$$

Let K values of the new variable y be drawn, $(y_1, \cdots, y_K)$ and let the spread

$$w = \sup (y_1, \cdots, y_K) - \inf (y_1, \cdots, y_K).$$

Fixing K, we consider the cumulative distribution function of w, $P_N(w)$, as $N \to \infty$. That is, we have K large samples of x and wish to examine the spread among their minima. It is evident intuitively that if $F(x) = 0$ for some finite x, these minima are bounded from below and will cluster near the vanishing point of $F(x)$, making $w \to 0$ statistically as $N \to \infty$. Our theorems also show that even when $y \to -\infty$ statistically, i.e., when $F(x) = 0$ for no finite x, the spread $w \to 0$ statistically if the tail of $F(x)$ is sufficiently small (e.g. Gaussian). On the other hand, if $F(x) = 0(e^{kx})$ as $x \to -\infty$, the distribution $P_N(w)$ does not peak as $N \to \infty$, while for larger tails (e.g. algebraic) $w \to +\infty$ statistically Two simple theorems are

I. *If*

$$\lim_{x \to -\infty} \frac{F(x)}{F(x+s)} = 1,$$

then

$$\lim_{N \to \infty} P_N(s) = 0.$$

II. *Let* $s > 0$. *If*

$$F(x_0) = 0 \text{ for some } x_0 > -\infty, \text{ or if}$$

$$\lim_{x \to -\infty} \frac{F(x)}{F(x+s)} = 0,$$

then

$$\lim_{N \to \infty} P_N(s) = 1.$$

Theorem I is directly applicable to distributions with algebraic tails, theorem II to Gaussian tails. We prove them both as corollaries of the more general results:

III. *If*

$$\liminf_{x \to -\infty} \frac{F(x)}{F(x+s)} = l$$

then

$$\limsup_{N \to \infty} P_N(s) \leq (1-l)^{K-1}.$$

IV. *Let* $s > 0$. If

$$F(x) = 0 \text{ for no finite } x \text{ and}$$

$$\limsup_{x \to -\infty} \frac{F(x)}{F(x+s)} = L,$$

then

$$\liminf_{N \to \infty} P_N(s) \geq [e^{-\alpha L} - e^{-\alpha}]^K$$

for any $\alpha > 0$.

Theorems III and IV together show that an exponential tail ($F(x) = 0(e^{kx})$) leads to a $P_N(w)$ which, asymptotically, is bounded away from 0 for any $w > 0$ and bounded away from 1 for w sufficiently small.

2. Proofs. Explicitly, for any $s \geq 0$,

$$P_N(s) = K \int_{-\infty}^{\infty} [G_N(x + s) - G_N(x)]^{K-1} \, dG_N(x + s). \tag{2}$$

Turning now to III: given $s > 0$, choose $x_1 = x_1(\epsilon)$ so that (i) $F(x_1) \neq 0$, and (ii), $x \leq x_1$ implies

$$\frac{F(x)}{F(x + s)} \geq l - \epsilon. \tag{3}$$

We then rewrite (2) as

$$P_N(s) = \int_{-\infty}^{x_1} \left[1 - \frac{G_N(x)}{G_N(x + s)}\right]^{K-1} dG_N(x + s)^K + \int_{x_1}^{\infty}. \tag{4}$$

Treating $G_N(x + s)^K$ as the independent variable, the first integral may be evaluated by the mean value theorem in the form

$$\left[1 - \frac{G_N(x_2)}{G_N(x_2 + s)}\right]^{K-1} \int_{-\infty}^{x_1} dG_N(x + s)^K \leq \left[1 - \frac{G_N(x_N)}{G_N(x_N + s)}\right]^{K-1} \tag{5}$$

with an appropriate $x_2 = x_2(N)$, $-\infty \leq x_2 \leq x_1$.

Using the form (2) of the integrand in the second term of (4), we may bound the latter by

$$K \int_{x_1}^{\infty} dG_N(x + s) \leq K[1 - G_N(x_1 + s)], \tag{6}$$

since

$$G_N(x + s) - G_N(x) \leq 1.$$

Now, by factoring (1),

$$\frac{G_N(x)}{G_N(x + s)} = \frac{F(x)}{F(x + s)} \frac{1 + Q + \cdots + Q^{N-1}}{1 + Q_s + \cdots + Q_s^{N-1}} \geq \frac{F(x)}{F(x + s)} \tag{7}$$

where $Q = 1 - F(x)$, $Q_s = 1 - F(x + s) \leq Q$. Combining (3), (4), (5), (6), and (7),

$$P_N(s) \leq [1 - l + \epsilon]^{K-1} + K[1 - G_N(x_1 + s)].$$

Since $F(x_1 + s) \geq F(x_1) > 0$, we have

$$\lim_{N \to \infty} G_N(x_1 + s) = 1.$$

Hence,

$$\limsup_{N \to \infty} P_N(s) \leq [1 - l + \epsilon]^{K-1}$$

and III follows by letting $\epsilon \to 0$. Then I follows immediately with $l = 1$, when we note that $P_N(s) \geq 0$.

To prove IV, choose any $\alpha > 0$. By hypothesis, for sufficiently large N we may always find $x_N = x_N(\alpha)$ such that

$$F(x_N) = \frac{\alpha L}{N}. \tag{8}$$

By hypothesis, and the monotonicity of $F(x)$, $x_N \to -\infty$ as $N \to \infty$. For any $\epsilon > 0$, therefore, we can find $N_0 = N_0(\alpha, \epsilon)$ such that $N \geq N_0$ implies

$$\frac{F(x_N)}{F(x_N + s)} \leq \frac{L}{1 - \epsilon} \tag{9}$$

or $F(x_N + s) \geq \frac{\alpha}{N}(1 - \epsilon)$. Directly from (2), since $s > 0$,

$$P_N(s) \geq K \int_{x_N - s}^{x_N} [G_N(x + s) - G_N(x)]^{K-1}\, dG_N(x + s)$$

$$\geq K \int_{x_N - s}^{x_N} [G_N(x + s) - G_N(x_N)]^{K-1}\, dG_N(x + s).$$

But this last integral is of the form

$$\int K(U - G)^{K-1}\, dU = (U - G)^K,$$

whence

$$P_N(s) \geq [G_N(x_N + s) - G_N(x_N)]^K,$$

or

$$P_N(s) \geq [(1 - F(x_N))^N - (1 - F(x_N + s))^N]^K. \tag{10}$$

By (8) and (9), therefore

$$P_N(s) \geq \left[\left(1 - \frac{\alpha L}{N}\right)^N - \left(1 - \frac{\alpha(1 - \epsilon)}{N}\right)^N\right]^K$$

Since this holds for all $N \geq N_0(\alpha, \epsilon)$,

$$\liminf_{N \to \infty} P_N(s) \geq [e^{-\alpha L} - e^{-\alpha(1-\epsilon)}]^K$$

This last, in turn, now holds for any $\epsilon > 0$, hence

$$\liminf_{N \to \infty} P_N(s) \geq [e^{-\alpha L} - e^{-\alpha}]^K.$$

This now holds for any $\alpha > 0$ Maximizing on α yields a sharper bound than the result of IV. The applicable part of II follows, when $L = 0$, by letting $\alpha \to \infty$. That the conclusion of II holds when $F(x_0) = 0$ for some finite x_0 follows from (10) with x_N replaced by some x_1 such that $F(x_1) = 0$, $F(x_1 + s) > 0$.

ON THE CONVERGENCE OF THE CLASSICAL ITERATIVE METHOD OF SOLVING LINEAR SIMULTANEOUS EQUATIONS[1]

By Edgar Reich

Massachusetts Institute of Technology

The classical iterative method, or Seidel method, is a scheme for solving the system of linear algebraic equations

$$\sum_{j=1}^{n} A_{ij}x_j = b_i, \qquad (i = 1, 2, \cdots, n),$$

by successive approximation, as follows:
If $x^{(\nu)} = (x_1^{(\nu)}, x_2^{(\nu)}, \cdots, x_n^{(\nu)})$ is the νth approximation of the solution, the $(\nu + 1)$st approximation, $x^{(\nu+1)} = (x_1^{(\nu+1)}, x_2^{(\nu+1)}, \cdots, x_n^{(\nu+1)})$, is obtained from the relations

$$\begin{cases} A_{11}x_1^{(\nu+1)} + A_{12}x_2^{(\nu)} + A_{13}x_3^{(\nu)} + \cdots + A_{1n}x_n^{(\nu)} = b_1, \\ A_{21}x_1^{(\nu+1)} + A_{22}x_2^{(\nu+1)} + A_{23}x_3^{(\nu)} + \cdots + A_{2n}x_n^{(\nu)} = b_2, \\ A_{31}x_1^{(\nu+1)} + A_{32}x_2^{(\nu+1)} + A_{33}x_3^{(\nu+1)} + \cdots + A_{3n}x_n^{(\nu)} = b_3, \\ \cdots \\ A_{n1}x_1^{(\nu+1)} + A_{n2}x_2^{(\nu+1)} + A_{n3}x_3^{(\nu+1)} + \cdots + A_{nn}x_n^{(\nu+1)} = b_n, \end{cases}$$

$x_1^{(\nu+1)}$ being obtained from the first equation, then $x_2^{(\nu+1)}$ from the second, and so on.

The given system can be written in matrix notation as $Ax = b$ where A is a non-singular square matrix of order n, and x and b are column vectors of order n. Let us define square matrices A_1 and A_2 as follows:

$$(A_1)_{ij} = \begin{cases} A_{ij} \text{ if } i \geq j \\ 0 \text{ if } i < j \end{cases},$$

$$(A_2)_{ij} = \begin{cases} A_{ij} \text{ if } i < j \\ 0 \text{ if } i \geq j \end{cases},$$

(Note that $A_1 + A_2 = A$.).

With this notation the Seidel method can be written as the matric difference equation

$$A_1x^{(\nu+1)} + A_2x^{(\nu)} = b.$$

Now various writers, among them C. E. Berry in this journal, (See list of refer-

[1] Work done under Office of Naval Research Contract N5ori60.

ences at end of this paper.) have shown that a necessary and sufficient condition for convergence, i.e., a necessary and sufficient condition for

$$\lim_{\nu \to \infty} (x_i^{(\nu)} - x_i) = 0, \qquad (i = 1, 2, \cdots, n),$$

is that

(1) A_1 has an inverse; that is $A_{11} \neq 0$ for any i.

(2) The characteristic roots of $(A_1^{-1}A_2)$ all have an absolute value smaller than unity.

It would be advantageous to rephrase the above condition, if possible, in terms of simpler requirements on A. As a step in this direction the following theorem is offered:

THEOREM. *If A is a real, symmetric nth-order matrix with all terms on its main diagonal positive, then a necessary and sufficient condition for all the n characteristic roots of $(A_1^{-1}A_2)$ to be smaller than unity in magnitude is that A is positive definite.*

PROOF. Let z_j be a characteristic vector of $(A_1^{-1}A_2)$ corresponding to the characteristic root μ_j. Then

$$(A_1^{-1}A_2)\, z_j = \mu_j z_j . \tag{1}$$

Premultiplying by $\bar{z}_i' A_1$, where the apostrophe and bar denote transposition and conjugation respectively:

$$\bar{z}_i' A_2 z_j = \mu_j \bar{z}_i' A_1 z_j . \tag{2}$$

Consider the bilinear form $\bar{z}_i' A z_j$.
We have

$$\bar{z}_i' A z_j = \bar{z}_i' A_1 z_j + \bar{z}_i' A_2 z_j = (1 + \mu_j)\, \bar{z}_i' A_1 z_j . \tag{3}$$

Interchanging i and j:

$$\bar{z}_j' A z_i = (1 + \mu_i) \bar{z}_j' A_1 z_i . \tag{4}$$

Taking the conjugate:

$$z_j' A \bar{z}_i = \bar{z}_i' A z_j = (1 + \bar{\mu}_i) z_j' A_1 \bar{z}_i = (1 + \bar{\mu}_i) \bar{z}_i' A_1' z_j . \tag{5}$$

Let D be the diagonal matrix with elements

$$D_{ij} = A_{ij}\delta_{ij} . \tag{6}$$

This makes $A_1' = D + A_2$.
Substituting this in (5):

$$\bar{z}_i' A z_j = (1 + \bar{\mu}_i)\, (\bar{z}_i' D z_j + \bar{z}_i' A_2 z_j) = (1 + \bar{\mu}_i) \bar{z}_i' D z_j + (1 + \bar{\mu}_i) \mu_j \bar{z}_i' A_1 z_j . \tag{7}$$

Eliminating $\bar{z}_i' A_1 z_j$ between relations (3) and (7) we obtain

$$(1 - \bar{\mu}_i \mu_j) \bar{z}_i' A z_j = (1 + \bar{\mu}_i)\, (1 + \mu_j) \bar{z}_i' D z_j . \tag{8}$$

To obtain the necessary condition we use the fact that we must have $|\mu_i| < 1$, and can therefore rewrite (8) as

$$(9) \qquad \bar{z}_i' A z_j = \frac{(1+\bar{\mu}_i)(1+\mu_j)}{1-\bar{\mu}_i\mu_j}\bar{z}_i' D z_j = \sum_{k=0}^{\infty}(1+\bar{\mu}_i)\bar{\mu}_i^k(1+\mu_j)\mu_j^k \bar{z}_i' D z_j.$$

If $x = \sum_{i=1}^{m} c_i z_i$ is any linear combination of the $m \leq n$ independent characteristic vectors of $(A_1^{-1}A_2)$ then

$$(10) \qquad \begin{aligned} \bar{x}' A x = \left(\sum_{i=1}^{m}\bar{c}_i\bar{z}_i'\right) A \left(\sum_{i=1}^{m} c_i z_i\right) &= \sum_{i,j=1}^{m}\bar{c}_i c_j \bar{z}_i' A z_j \\ &= \sum_{i,j=1}^{m}\bar{c}_i c_j \sum_{k=0}^{\infty}(1+\bar{\mu}_i)\bar{\mu}_i^k(1+\mu_j)\mu_j^k \bar{z}_i' D z_j, \end{aligned}$$

or

$$\bar{x}' A x = \sum_{k=0}^{\infty}\bar{y}_k' D y_k$$

where

$$y_k = \sum_{i=1}^{m} c_i(1+\mu_i)\mu_i^k z_i.$$

Since by hypothesis $A_{11} > 0$, D is evidently positive definite, and therefore

$$(11) \qquad \bar{x}' A x > 0.$$

In case the characteristic roots μ_i, $(i = 1, 2, \cdots n)$, are all distinct there will be n independent z_i assured, and in that case (11) implies that A is positive definite Consider, on the other hand, the case where the μ_i are not all distinct. Note that (a) the definiteness properties of a matrix are not changed by sufficiently small alterations in the elements; (b) the μ's depend continuously on the elements of A; (c) the discriminant of (1) is a polynomial in the A_{ij} that does not vanish identically.[2] It follows that A must be positive definite even in the case of repeated roots because an arbitrarily small change in A will separate any multiple μ's, still keeping them smaller than unity in magnitude, and not changing the definiteness properties of A.

This completes the proof that the condition given in the statement of the theorem is necessary. Now to prove sufficiency:

Setting $i = j$ in relation (8) we obtain

$$(12) \qquad (1 - |\mu_i|^2)\bar{z}_i' A z_i = |1 + \mu_i|^2 \bar{z}_i' D z_i$$

Since both A and D are positive definite

$$(13) \qquad \bar{z}_i' A z_i > 0 \text{ and } \bar{z}_i' D z_i > 0.$$

[2] The fact that the discriminant is not identically zero follows from easily constructible counter-examples.

Moreover, we cannot have $\mu_i = -1$ because that would mean by (3) that

$$0 = \bar{z}_i' A_1 z_i + \bar{z}_i' A_2 z_i = \bar{z}_i' A z_i .$$

Relation (12) thus implies

$$1 - |\mu_i|^2 > 0 \tag{14}$$

i.e. $|\mu_i| < 1$ as was to be proved.

The part of the theorem giving the sufficient condition was already obtained by L. Seidel [1] and G. Temple in a somewhat more indirect fashion.

REFERENCES

[1] L. Seidel, "Über ein Verfahren die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineare Gleichungen überhaupt, durch successive Annäherung aufzulösen," *Abhandlungen der Mathematisch-Physikalischen Classe der Königlich Bayerischen Akademie der Wissenschaften*, Vol. 11 (1874), pp 81–108.

[2] C E Berry, "A criterion of convergence for the classical iterative method of solving linear simultaneous equations," *Annals of Math. Stat*, Vol 16 (1945), pp 398–400.

[3] L Cesari, "Sulla risoluzione dei sistemi di equazioni lineari per approssimazioni successive," *Rassegna delle Poste, dei Telegrafi e dei Telefoni*, Anno 9 (1931)

[4] L. Cesari, "Sulla risoluzione dei sistemi di equazioni lineari per approssimazioni successive," *Reale Accademia Nazionale dei Lincei, Serie 6, Classe di Szienze fisiche, matematiche e naturali, Rendiconti*, Vol 25 (1937), pp. 422–428

[5] J Morris, *The Escalator Method in Engineering Vibration Problems*, Chapman and Hall Ltd., London 1947, pp. 63–70

[6] R J Schmidt, "On the numerical solution of linear simultaneous equations by an iterative method," *Phil. Mag*, Ser 7, Vol. 32 (1941), pp. 369–383.

SOME RECURRENCE FORMULAE IN THE INCOMPLETE BETA FUNCTION RATIO

By T. A. Bancroft

Alabama Polytechnic Institute

1. Introduction. It is well known that the incomplete beta function ratio, defined by

$$I_x(p, q) = \frac{B_x(p, q)}{B(p, q)}, \tag{1}$$

where

$$B_x(p, q) = \int_0^x x^{p-1}(1 - x)^{q-1}\, dx, \tag{2}$$

and

$$B(p, q) = B_1(p, q), \tag{3}$$

is of importance in probability distribution theory, and, hence, also in obtaining exact probability values in making tests of statistical hypotheses. In constructing certain extensions [1] of Karl Pearson's "Tables of the Incomplete Beta-Function" [2], the recurrence formulae contained in the following sections were derived.

2. Derivation of formulae. The incomplete beta function, $B_x(p, q)$ may be considered as a special case of the hypergeometric series, $F(a, b, c, x)$, thus

$$B_x(p, q) = \frac{x^p}{p} F(p, 1 - q, p + 1, x). \tag{4}$$

The series converges for $|x| \leq 1$, if and only if $a + b < c$. By setting $a = p$, $b = 1 - q$, and $c = p + 1$, as in (4), all conditions are satisfied, if we also take $q > 0$.

Recurrence formulae for $F(a, b, c, x)$, e. g., in the work of Magnus and Oberhettinger [3], may now be directly converted for use with $B_x(p, q)$ or $I_x(p, q)$. In particular, using the three identities on page 9 of [3], with x replacing z, we have

$$cF(a, b, c, x) + (b - c)F(a + 1, b, c + 1, x) - b(1 - x)F(a + 1, b + 1, c + 1, x) = 0, \tag{5}$$

$$c(c - ax - b)F(a, b, c, x) - c(c - b)F(a, b - 1, c, x) + abx(1 - x)F(a + 1, b + 1, c + 1, x) = 0, \tag{6}$$

$$cF(a, b, c, x) - cF(a, b + 1, c, x) + axF(a + 1, b + 1, c + 1, x) = 0, \tag{7}$$

with $a = p$, $b = 1 - q$, and $c = p + 1$, we obtain in turn

$$xI_x(p, q) - I_x(p + 1, q) + (1 - x)I_x(p + 1, q - 1) = 0 \tag{8}$$

$$(p + q - px)I_x(p, q) - qI_x(p, q + 1) - p(1 - x)I_x(p + 1, q - 1) = 0 \tag{9}$$

$$qI_x(p, q + 1) + pI_x(p + 1, q) - (p + q)I_x(p, q) = 0. \tag{10}$$

Formula (8) is the basic recurrence formula used in the construction of Karl Pearson's [2] tables. Formula (10) was obtained, incidentally, by the author [4] in a different connection and manner.

Formulae (8), (9), and (10) may now be combined to give other useful formulae, e. g.,

$$qI_x(p + 1, q + 1) + (\overline{p + q}x - q)I_x(p + 1, q) - (p + q)xI_x(p, q) = 0, \tag{11}$$

$$pI_x(p + 1, q + 1) + (q - \overline{p + q}x)I_x(p, q + 1) - (p + q)(1 - x)I_x(p, q) = 0, \tag{12}$$

$$(13)\quad (p+q-1)xI_x(p-1,q) - (\overline{p+q-1}x+p)I_x(p,q) + pI_x(p+1,q) = 0,$$

$$(14)\quad (p+q)(1-x)I_x(p+1,q-1) - \{(p+q)(1-x)+q\}I_x(p+1,q) + pI_x(p+1,q+1) = 0.$$

Notice that the sum of the coefficients is always zero

By a repeated use of (10) it is possible to obtain the formulae

$$(15)\quad I_x(p+n,q) = \frac{1}{(p+n-1)^{(n)}}\sum_{r=0}^{n}(-1)^r \cdot \binom{n}{r}(p+q+n-1)^{(n-r)}(q+r-1)^{(r)}I_x(p,q+r),$$

$$(16)\quad I_x(p,q+n) = \frac{1}{(q+n-1)^{(n)}}\sum_{r=0}^{n}(-1)^r \cdot \binom{n}{r}(p+q+n-1)^{(n-r)}(p+r-1)^{(r)}I_x(p+r,q),$$

where $(p+q+n-1)^{(n-r)}$, etc., refer to the factorial notation, e. g.,

$$[p+q+(n-1)]^{(n-r)} = (p+q+n-1)(p+q+n-2)\cdots(p+q+r).$$

3. An application. Formulae (15) and (16) may be used to write general formulae for obtaining values of $I_x(p,q)$ where p or q may be greater than 50, i. e., for such values outside the range of Karl Pearson's tables. In particular,

$$(17)\quad I_x(50+n,q) = \frac{1}{49+n)^{(n)}}\left[n+q+49)^{(n)}I_x(50,q) - \binom{n}{1}q(n+q+49)^{(n-1)}I_x(50,q+1)\cdots(-1)^n(q+n-1)^{(n)}I_x(50,q+n)\right]$$

and

$$(18)\quad I_x(p,50+n = \frac{1}{(49+n)^{(n)}}\left[(n+p+49)^{(n)}I_x(p,50) - \binom{n}{1}p(n+p+49)^{(n-1)}I_x(p+1,50)\cdots(-1)^n(p+n-1)^{(n)}I_x(p+n,50)\right].$$

It should be noted for (17) that as n increases the range of values that can be obtained outside Karl Pearson's tables are reduced since the last term of (17) contains $I_x(50,q+n)$. A similar observation is noted for (18). From a practical standpoint the computational labor restricts n to fairly small values. Using (17) we may easily compute for example,

$$I_{.60}(52,48) = I_{.60}(50+2,48)$$

$$= \frac{1}{(51)(50)}[(99)(98)I_{.60}(50,48) - 2(99)(48)I_{.60}(50,49) + (49)(48)I_{.60}(50,50)].$$

Substituting the necessary values from Karl Pearson's tables we calculate

$$I_{.60}(52, 48) = .9465248.$$

Similarly using (18) we may calculate

$$I_{.40}(48, 52) = .0534752.$$

As a check on the computations, we use the well-known identity

$$I_x(p, q) = 1 - I_{1-x}(p', q'),$$

where $p' = q$ and $q' = p$. Then

$$\begin{aligned} I_{.40}(48, 52) &= 1 - I_{.60}(52, 48) \\ &= 1 - .9465248 \\ &= .0534752. \end{aligned}$$

In like manner formulae (15) and (16) may be used to write general formulae for obtaining half values for p or q greater than 10.5, i. e., for values not included in Karl Pearson's tables. In particular,

$$\begin{aligned} I_x(10.5 + n, q) = \frac{1}{(9.5 + n)^{(n)}} \Big[&(9.5 + q + n)^{(n)} I_x(10.5, q) - \binom{n}{1} \\ &\cdot q(9.5 + q + n)^{(n-1)} I_x(10.5, q + 1) \cdots (-1)^n (q + n - 1)^{(n)} I_x(10.5, q + n) \Big], \end{aligned} \tag{19}$$

and

$$\begin{aligned} I_x(p, 10.5 + n) = \frac{1}{(9.5 + n)^{(n)}} \Big[&(9.5 + p + n)^{(n)} I_x(p, 10.5) - \binom{n}{1} \\ &\cdot p(9.5 + p + n)^{(n-1)} I_x(p + 1, 10.5) \cdots (-1)^n (p + n - 1)^{(n)} I_x(p + n, 10.5) \Big]. \end{aligned} \tag{20}$$

Using (19) we may compute

$$\begin{aligned} I_{.60}(12.5, 8) = \frac{1}{(11.5)^{(2)}} [&(19.5)^{(2)} I_{.60}(10.5, 8) - 2(8)(19.5) I_{.60}(10.5, 9) \\ &+ (9)(8) I_{.60}(10.5, 10)], = .4512367. \end{aligned}$$

Similarly using (20) we obtain

$$I_{.40}(8, 12.5) = .5487633.$$

Employing the check formula,

$$\begin{aligned} I_{.40}(8, 12.5) &= 1 - I_{.60}(12.5, 8) \\ &= 1 - .4512367 \\ &= .5487633. \end{aligned}$$

Thanks are due to Dr. J. C. P. Miller, Technical Director, Scientific Computing Service, Limited, London, England, for helpful suggestions in the preparation of this paper.

REFERENCES

[1] T. A. Bancroft, "Some extensions of the incomplete beta function tables" (in preparation)

[2] Karl Pearson, *Tables of the Incomplete Beta-Function*, Cambridge University Press, 1934.

[3] Wilhelm Magnus und Fritz Oberhettinger, *Formeln and Sätze fur die Speziellen Funktionen der Mathematischen Physik*, Julius Springer, Berlin, 1943.

[4] T A Bancroft, "On biases in estimation due to the use of preliminary tests of significance, *Annals of Math. Stat*," Vol 15 (1944).

ON A THEOREM BY WALD AND WOLFOWITZ

By Gottfried E. Noether

New York University

Let $\mathfrak{H}_n = (h_1, \cdots, h_n)$, $(n = 1, 2, \cdots)$, be sequences of real numbers and for all n denote by $H_{e_1 \cdots e_m}$ the symmetrical function generated by $h_1^{e_1} \cdots h_m^{e_m}$, i e, $H_{e_1 \cdots e_m} = \Sigma h_{i_1}^{e_1} \cdots h_{i_m}^{e_m}$ where the summation is extended over the $n(n-1) \cdots (n-m+1)$ possible arrangements of the m integers $i_1, \cdots, i_m$, such that $1 \leq i_j \leq n$ and $i_j \neq i_k$, $(j, k = 1, \cdots, m)$. According to Wald and Wolfowitz [1] the sequences $\mathfrak{H}_n$ are said to satisfy condition W, if for all integral $r > 2$

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(h_i - \bar{h})^r}{\left[\frac{1}{n}\sum_{i=1}^{n}(h_i - \bar{h})^2\right]^{r/2}} = O(1),^1$$

where $\bar{h} = 1/n \, \Sigma_{i=1}^{n} h_i$.

Given sequences $\mathfrak{A}_n = (a_1, \cdots, a_n)$ and $\mathfrak{D}_n = (d_1, \cdots, d_n)$, consider the chance variable

$$L_n = d_1 x_1 + \cdots + d_n x_n,$$

where the domain of $(x_1, \cdots, x_n)$ consists of the $n!$ equally likely permutations of the elements of $\mathfrak{A}_n$. Then it is shown in [1] that if the sequences $\mathfrak{A}_n$ and $\mathfrak{D}_n$ satisfy condition W, the distribution of $L_n^0 = (L_n - EL_n)/\sigma(L_n)$ approaches the normal distribution with mean 0 and variance 1 as $n \to \infty$. These conditions

[1] The symbol O, as well as the symbols o and $\sim$ to be used later, have their usual meaning. See e g Cramér [2, p. 122]

for asymptotic normality can be weakened. It will be shown that the following theorem holds:

THEOREM. *L_n^0 is asymptotically normal with mean 0 and variance 1 provided the sequences $\mathfrak{D}_n$ satisfy condition W while for the sequences $\mathfrak{A}_n$*

$$\frac{\sum_{i=1}^{n}(a_i - \bar{a})^r}{\left[\sum_{i=1}^{n}(a_i - \bar{a})^2\right]^{r/2}} = o(1), \qquad (r = 3, 4, \cdots). \tag{1}$$

We note that L_n^0 is [illegible] is replaced by $[1/n\Sigma_{i=1}^{n}(a_i - \bar{a})^2]^{-1/2}(a_i - \bar{a})$ and d_i by $[1/n\ \Sigma\ \ldots\ \ldots\ \ldots\ - \bar{d})$. Therefore it is sufficient to prove asymptotic normality provided

$$D_1 = 0, \quad D_2 = n, \quad D_r = O(n), \qquad (r = 3, 4, \cdots); \tag{2}$$

$$A_1 = 0, \quad A_2 = n, \quad A_r = o(n^{r/2}), \qquad (r = 3, 4, \cdots). \tag{3}$$

Then

$$EL_n = D_1 Ex_1 = 0,$$

$$\operatorname{var} L_n = EL_n^2 = D_2 Ex_1^2 + D_{11}Ex_1x_2$$

$$= \frac{1}{n}A_2D_2 + \frac{1}{n(n-1)}(A_1^2 - A_2)(D_1^2 - D_2) \sim n,$$

and it is sufficient to show that $n^{-r/2}EL_n^r$ tends to the rth moment of a normal distribution with mean 0 and variance 1.

Now we can write

$$\mu_r = n^{-r/2}EL_n^r = n^{r/2}\sum_{i_1=1}^{n}\cdots\sum_{i_r=1}^{n}Ed_{i_1}x_{i_1}\cdots d_{i_r}x_{i_r}$$

$$= n^{-r/2}[D_rEx_1^r + \cdots + c(r, e_1, \cdots, e_m)D_{e_1\cdots e_m}Ex_1^{e_1}\cdots x_m^{e_m} \tag{4}$$

$$+ \cdots + D_{1\cdots1}Ex_1\cdots x_r]$$

where $e_1 + \cdots + e_m = r$ with e_k, $(k = 1, \cdots, m)$, positive integral and the coefficient $c(r, e_1, \cdots, e_m)$ stands for the number of ways in which the r indices $i_1, \cdots, i_r$ can be tied in m groups of size $e_1, \cdots, e_m$, respectively, so as to produce the terms of $D_{e_1\cdots e_m}Ex_1^{e_1}\cdots x_m^{e_m}$.

Since $Ex_1^{e_1}\cdots x_m^{e_m} \sim n^{-m}A_{e_1\cdots e_m}$ we have

$$n^{-r/2}D_{e_1\cdots e_m}Ex_1^{e_1}\cdots x_m^{e_m} \sim n^{-(r/2+m)}D_{e_1\cdots e_m}A_{e_1\cdots e_m} = B(r, e_1, \cdots, e_m), \text{ say.} \tag{5}$$

LEMMA. *$B(r, e_1, \cdots, e_m) \sim 0$ unless*

$$m = r/2, \qquad e_1 = \cdots = e_{r/2} = 2. \tag{6}$$

In that case $B(r, 2, \cdots, 2) \sim 1$.

Before proving this lemma we shall show that our theorem follows immediately By (4) μ_r is the sum of a finite number of expressions $B(r, e_1, \cdots, e_m)$.

Therefore if $r = 2s + 1$, $(s = 1, 2, \cdots)$, $\mu_{2s+1} \sim 0$, since at least one of the e_k, $(k = 1, \cdots, m)$, in all the $B(2s + 1, e_1, \cdots, e_m)$ adding up to μ_{2s+1} must be odd. If $r = 2s$, $\mu_{2s} \sim c(2s, 2, \cdots, 2)$. Since the first index in (4) can be tied with any one of the other $2s - 1$ indices, the next free index with any one of the remaining $2s - 3$ indices, etc., it is seen that $\mu_{2s} \sim (2s - 1)(2s - 3) \cdots 3$. However these are the moments of a normal distribution with mean 0 and variance 1 This proves the theorem.

PROOF OF LEMMA. Define $A(j_1, \cdots, j_h) = A_{j_1} \cdots A_{j_h}$ Then $A_{e_1 \cdots e_m}$ is the sum of a finite number of expressions $A(j_1, \cdots, j_h)$, where the j_g, $(g = 1, \cdots, h)$, are obtained from $e_1, \cdots, e_m$ by addition in such a way that

$$j_1 + \cdots + j_h = e_1 + \cdots + e_m = r \tag{7}$$

Since by (3) $A_1 = 0$, we need only consider those $A(j_1, \cdots, j_h)$ for which $j_g \geq 2$, $(g = 1, \cdots, h)$ If some $j_g > 2$ by (3) and (7)

$$A(j_1, \cdots, j_h) = o(n^{r/2}). \tag{8}$$

If $j_g \equiv 2$,

$$A(2, \cdots, 2) = A_2^{r/2} = n^{r/2}. \tag{9}$$

This last case can only happen if r is even and e_k, $(k = 1, \cdots, m)$, equals either 1 or 2. Therefore, unless (6) is true

$$m > r/2. \tag{10}$$

Similarly, writing $D_{e_1 \cdots e_m}$ as a sum of products of the kind $D_{j_1} \cdots D_{j_h}$ it is seen that by (2)

$$D_{e_1 \cdots e_m} = \begin{cases} O(n^m) & \text{if } m < r/2 \\ O(n^{r/2}) & \text{if } m \geq r/2. \end{cases} \tag{11}$$

Thus by (8)–(11)

$$A_{e_1 \cdots e_m} D_{e_1 \cdots e_m} = o(n^{r/2+m}), \tag{12}$$

unless (6) is true. In that case

$$A_{2 \cdots 2} \sim A_2^{r/2} = n^{r/2}, \tag{13}$$

$$D_{2 \cdots 2} \sim D_2^{r/2} = n^{r/2}. \tag{14}$$

(12)–(14) together with (5) prove the lemma

Let $a_1, a_2, \cdots$ be independent observations on the same chance variable Y. We may ask what conditions have to be imposed on the distribution of Y to insure—at least with probability 1—that condition (1) is satisfied. Wald and Wolfowitz state in Corollary 2 of [1] that provided Y has positive variance and finite moments of all orders the $a_1, a_2, \cdots$ satisfy condition W with probability 1 and therefore insure asymptotic normality of L_n provided the sequences $\mathfrak{D}_n$ satisfy condition W. On the other hand, it can be shown that the $a_1, a_2, \cdots$

satisfy condition (1) with probability 1, provided Y has positive variance and a finite absolute moment of order 3. Thus condition (1) constitutes a considerable improvement over condition W.

REFERENCES

[1] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 358-372.
[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton, 1946.

ON SUMS OF SYMMETRICALLY TRUNCATED NORMAL RANDOM VARIABLES

BY Z. W. BIRNBAUM AND F. C. ANDREWS[1]
University of Washington, Seattle

1. Introduction. Let X_a be the random variable with the probability density

$$f_a(x) = \begin{cases} Ce^{-x^2/2} & \text{for } |x| \leq a \\ 0 & \text{for } |x| > a, \end{cases} \tag{1.1}$$

obtained from the normal probability density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ by symmetrical truncation at the "terminus" $|x| = a$, and let $S_a^{(m)}$ be the sum of m independent sample-values of X_a. We consider the following *problem*: An integer $m \geq 2$ and the real numbers $A > 0$, $\epsilon > 0$ are given; how does one have to choose the terminus a so that the probability of $|S_a^{(m)}| \geq A$ is equal to ϵ,

$$P(|S_a^{(m)}| \geq A) = \epsilon? \tag{1.2}$$

This problem arises for example when single components of a product are manufactured under statistical quality control, so that each component has the length $Z = k + X$ where X has the probability density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, and the final product consists of m components so that its total length S is the sum of the lengths of the components. We wish to have probability $1 - \epsilon$ that S differs from mk by not more than a given A. To achieve this we decide to reject each single component for which $|Z - k| = |X| > a$; how do we determine a?

The exact solution of this problem would require laborious computations.[2] In the present paper methods are given for obtaining approximate values of a which are "safe", that is such that

$$P(|S_a^{(m)}| \geq A) \leq \epsilon. \tag{1.3}$$

[1] Research done under the sponsorship of the Office of Naval Research

[2] A similar problem has been studied by V. J. Francis [2] for one-sided truncation, he actually had the exact probabilities for the solution of his problem computed and tabulated for $m = 2, 4$.

In deriving these safe values, use will be made of theorems on random variables with comparable peakedness, for which the reader is referred to a previous paper [1].

2. The safe value a_1. For fixed $a > 0$, we consider the normal random variable Y_a with expectation 0 and with probability density $g_a(Y_a)$ such that $g_a(0) = f_a(0)$. It is easily seen that Y_a has the standard deviation

$$\sigma_a = \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-t^2/2}\, dt, \tag{2.1}$$

and that $g_a(\xi) < f_a(\xi)$ for $|\xi| \leq a$, $g_a(\xi) > 0 = f_a(\xi)$ for $|\xi| > a$. Hence, applying Theorem 1 in [1], we conclude that

$$P(|S_a^{(m)}| \geq A) \leq \frac{2}{\sqrt{2\pi}} \int_{(A/\sigma_a\sqrt{m})}^{\infty} e^{-t^2/2}\, dt. \tag{2.2}$$

If m, A, and ϵ are given, we determine ξ_ϵ from tables of the normal probability integral so that $\frac{2}{\sqrt{2\pi}} \int_{\xi_\epsilon}^{\infty} e^{-t^2/2}\, dt = \epsilon$, set $\sigma_a = \frac{A}{\xi_\epsilon\sqrt{m}}$ in (2.1), and solve the equation

$$\frac{A}{\xi_\epsilon\sqrt{m}} = \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-t^2/2}\, dt \tag{2.3}$$

for a using again tables of the normal probability integral. In view of (2.2) this solution satisfies (1.3) and hence is safe; it will be denoted by a_1.

3. The safe value a_2. A direct application of Theorem 2 in [1] yields the inequality

$$P(|S_a^{(m)}| \geq A) \leq \frac{1}{2^{m-1}m!} \sum_{\frac{1}{2}(m+A/a) < j \leq m} (-1)^j \binom{m}{j} \left(\frac{A}{a} + m - 2j\right)^m = h_m\left(\frac{A}{a}\right) \tag{3.1}$$

for $0 \leq A \leq ma$. Hence by equating $h_m(A/a)$ to ϵ and solving for a, we obtain a safe value which will be denoted by a_2. It is of interest to note that (3.1) *is true* not only for $f_a(x)$ defined by (1.1) i.e. truncated normal, but *for any probability density $f_a(x)$ which is symmetrical and unimodal*, since these are the only assumptions needed for Theorem 2 in [1].

4. Solution for large m. The random variable X_a has the variance

$$\sigma^2(X_a) = 1 + \frac{2\phi''(a)}{2\phi(a) - 1} \tag{4.1}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Hence, according to the central limit theorem, we have the approximate equality

$$P(|S_a^{(m)}| \geq A) = \frac{2}{\sqrt{2\pi}} \int_{(A/\sigma(X_a)\sqrt{m})}^{\infty} e^{-t^2/2}\, dt \tag{4.2}$$

for m sufficiently large.

It can be reasonably expected that the cumulative distribution of $S_a^{(m)}$ differs from its limiting normal probability integral by less than the cumulative distribution of the sum $U_a^{(m)}$ of m independent uniform variables in $(-a, +a)$ differs from its limiting normal probability integral. Already for $m = 4$ the cumulative distribution of $U_a^{(m)}$ differs from the corresponding normal cumulative by less than .0075. Equally good or better approximation may, therefore, be expected for the distribution of $S_a^{(m)}$, so that the error in the approximate equality (4.2) between the two-tail probabilities should be less than .015 for $m = 4$, and still less for $m > 4$.

Equating the right-hand term of (4.2) to ϵ and solving for $\sigma^2(X_a)$, we obtain

$$\sigma^2(X_a) = 1 + \frac{2\phi''(a)}{2\phi(a) - 1} = \frac{1}{m}\left(\frac{A}{\xi_\epsilon}\right)^2,$$

an equation which can be solved for a with the aid of tables of $\phi(x)$ and $\phi''(x)$. We denote this value of a by α_1.

5. Use of the different solutions in practice. From the foregoing it appears that the following procedure may be followed in solving our problem in any definite case:

If m is large, α_1 is very close to the exact solution of (1.3) and may be used safely.

If m is not large but $m \geq 5$, it is conjectured that α_1 is such that the left-hand term in (1.3), for $a = \alpha_1$, differs from ϵ by less than 0.015.

If $m \leq 4$, the larger of a_1 and a_2 should be used. Table I contains the A for which a_1 and a_2 have the same value, say a'; a_1 or a_2 should be used if the given A is greater or smaller, respectively, than the tabulated value. The value a_1 is easily computed from a table of the normal probability integral by the procedure of section 2 The value a_2 can be obtained by reading off A/a_2 from Table II.

TABLE I

Values of A for which $a_1 = a_2 = a'$ for given m, ϵ

ϵ \ m	2		3		4	
	A	a'	A	a'	A	a'
.001	4.568	2 357	5.446	2.008	6.152	1 842
.002	4.258	2 228	5 059	1.918	5.717	1.779
005	3 808	2.047	4.512	1.799	5.111	1.697
01	3.438	1 910	4.074	1.712	4.632	1.640
.02	3.034	1.765	3 614	1.630	4.131	1.589
.05	2 456	1.581	2.970	1.533	3.425	1 529

TABLE II

Values of A/a_2 for given m, ϵ

ϵ \ m	2	3	4
.001	1.937	2.712	3 339
.002	1 911	2.637	3.213
.005	1.859	2 507	3 011
.01	1.800	2.379	2.824
.02	1.718	2.217	2.600
.05	1 553	1.937	2.240

6. Examples. 1) $A = 3.8$, $m = 4$, $\epsilon = .05$ Since A is greater than the value 3.425 in Table I, we compute $a_1 = 2.162$. From Table II we would obtain $A/a_2 = 2.240$ and thus $a_2 = 1.696 < a_1$. 2) $A = 3, m = 4, \epsilon = .02$. Since $A < 4$ 131, we read $A/a_2 = 2.600$ from Table II and obtain $a_2 = 1.153$ which will be greater than a_1. 3) $A = 5$, $m = 30$, $\epsilon = .05$. Using the method of section 4 we obtain $\alpha_1 = 1$ 62.

REFERENCES

[1] Z W Birnbaum, "On random variables with comparable peakedness," *Annals of Math Stat.*, Vol 19 (1948), pp 76–81

[2] V J Francis, "On the distribution of the sum of n sample values drawn from a truncated normal population," *Roy Stat. Soc Jour Suppl*, Vol 8 (1946), pp 223–232

A CERTAIN CUMULATIVE PROBABILITY FUNCTION

By Sister Mary Agnes Hatke, O.S.F.

St. Francis College, Ft. Wayne, Indiana

Graduations of empirically observed distributions show that the cumulative probability function $F(x) = 1 - (1 + x^{1/c})^{-1/k}$ is a practical tool for fitting a smooth curve to observed data. The graduations are comparable with those obtained by the Pearson system, Charlier, and others and are accomplished with simple calculations. Given distributions are graduated by the method of moments Theoretical frequencies are obtained by evaluation of consecutive values of $F(x)$ by use of calculating machines and logarithms, and by differencing $NF(x)$. No integration nor heavy interpolation is involved, such as may be required in graduation by a classical frequency function Burr [1] constructed tables of ν_1, σ, α_3, and α_4 values for the function $F(x)$ for certain combinations of integral values of $1/c$ and $1/k$. In these tables curvilinear interpolation must be used in finding an $F(x)$ with desired moments. The writer constructed more extensive tables for the same cumulative function with c and k a variety of real positive numbers less than or equal to one, such that linear interpolation can be used to determine the parameters c and k for an $F(x)$ that has α_3 and α_4 approximately the same as those of the distribution to be graduated. These tables have been deposited with Brown University. Microfilm or photostat copies may be obtained upon request to the Brown University Library.

The writer used the definitions of cumulative moments and the formulas for the ordinary moments ν_1, σ, α_3, and α_4 in terms of cumulative moments as developed by Burr. These latter moments were tabulated for the function $F(x)$ having various combinations of parameters c and k, c ranging from 0.050 to 0 675 and k from 0 050 to 1.000, each at intervals of 0.025 Within these ranges only those combinations of c and k were used which yielded α_3 of approximately 1 or less and α_4 values of 6 or less, since such moments are most common in practice.

It can be verified that over most of the area of the table α_3 values obtained

by linear and by curvilinear interpolation on k (or on c) differ by less than 0.001 and values of α_4 by approximately 0.01 or less. If α_3 = constant and α_4 = constant curves are plotted on c, k axes, it will be seen that there exists only one solution (c, k) of the equations $\alpha_3 = B(c, k)$ and $\alpha_4 = C(c, k)$. Furthermore, some α_4 curves intersect two α_3 curves representing the same $|\alpha_3|$. Thus the chance of finding an appropriate function $F(x)$ for graduation is increased since by reversal of scale an $F(x)$ with a positive α_3 may be used to graduate a distribution with a negative α_3, and conversely.

Graduation of an observed frequency distribution is easily accomplished. Linear interpolation on k for a fixed c seems to be the best method for determining

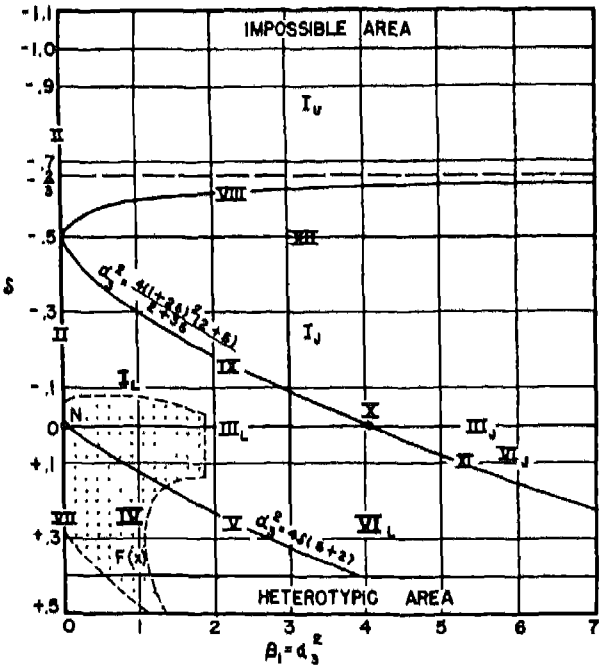

FIG. 1. The α_3^2, δ chart for the Pearson system of frequency curves and the area covered by $f(x) = 1 - (1 + x^{1/c})^{-1/k}$ (subscript L = bell-shaped)

the parameters of an $F(x)$ that has α_3 exactly the same and α_4 nearly the same as the observed α_3 and α_4. If the observed α_3 and α_4 are fairly close to an entry in the table, no interpolation is required. Direct linear interpolation is used to determine ν_1 and σ for the c and k just found. Letting M and S be the mean and standard deviation of the given distribution, the formula,

$$\frac{x - \nu_1}{\sigma} = t = \frac{X - M}{S}$$

is used to translate the class limits X of the given distribution to the corresponding x's of $F(x)$. For any x that is negative the quantity $1 + x^{1/c}$ is taken as one

to make $F(-x) = 0$ in accordance with the definition of $F(x)$ [1]. The values of $(1 + x^{1/c})^{-1/k}$ for the various x's are computed by logarithms and differenced to obtain the probabilities for the given class intervals, according to equation

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx = F(b) - F(a).$$

The respective theoretical frequencies are these probabilities multiplied by N, the number of cases

The headings that proved satisfactory for the columns of the graduation work-sheet are: class intervals (in observed physical units), X (u if unit class-interval is used), f_{obs}, x, $1 + x^{1/c}$, $N/(1 + x^{1/c})^{1/k}$, and f_{th}.

The relation of $F(x)$ to the Pearson system of frequency curves is presented in Figure 1, which is a reproduction of a major part of Craig's chart for α_3^2 and δ [2]. In this chart the parameters of the twelve Pearson curves are expressed in terms of α_3^2 and δ, where $\delta = (2\alpha_4 - 3\alpha_3^2 - 6)/(\alpha_4 + 3)$. Values of α_3^2 and δ were computed for $F(x) = 1 - (1 + x^{1/c})^{-1/k}$ in which c and k were assigned the values listed in the α_3, α_4 table. The dotted area superimposed on the Craig chart is that covered by these α_3^2, δ values for $F(x)$ Although it is small in size compared to the total area, it contains a part of the areas representing the three main Pearson curves, I, IV, and VI, as well as the point for the normal curve and part of the line on which lie the points corresponding to the bell-shaped curves of the Type III functions. It also includes transitional Types V and VII. Thus the function $F(x)$ covers part of an important area on the α_3^2, δ chart for the Pearson curves.

The function $F(x)$ was used to graduate satisfactorily several observed distributions classified as Pearson types, including the three main Types, I, IV, and VI, and transitional Types III and VII.

One advantage in the use of this cumulative function $F(x)$ is that it takes but one symbolic form with the area covered, whereas the Pearson-system curves require several different expressions of various complexity requiring identification of type Furthermore, graduation by a Pearson function generally involves approximate integration or heavy interpolation in the incomplete beta function tables for the evaluation of the integrals of the Pearson functions, whereas graduation by a function $F(x)$ is easily and quickly performed since $F(x)$ only involves two number-parameters readily determined by means of the α_3, α_4 table and straight arithmetic.

The writer is deeply indebted to Professor Irving W. Burr of Purdue University for valuable suggestions in this study.

REFERENCES

[1] I. W. Burr, "Cumulative frequency functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 215–232.

[2] C. C. Craig, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol 7 (1936), pp 16–28

ABSTRACTS OF PAPERS

(Presented at the Berkeley Meeting of the Institute, June 16-18, 1949)

1. Extension of a Theorem of Blackwell.

E. W. Barankin, University of California, Berkeley.

It is proved that Blackwell's method of uniformly improving the variance of an unbiased estimate by taking the conditional expectation with respect to a sufficient statistic, is, in fact, similarly effective on every absolute central moment of order $s \geqq 1$. The method leads to finer detail concerning the relationship between an estimate and its thus derived one. (This paper was prepared with the partial support of the Office of Naval Research.)

2. On the Existence of Consistent Tests.

Agnes Berger, Columbia University., New York.

Let $\mathfrak{M}(\mathfrak{B})$ denote the space of all probability-measures defined over a common Borel-field $\mathfrak{B}$. Let $\{m\} = M$, $\{m'\} = M'$ be two disjoint subsets of $\mathfrak{M}(\mathfrak{B})$ and let H_0 (H_1) be the hypothesis stating that the unknown distribution is in M (M'). In Neyman's terminology H_0 can be consistently tested against H_1 if to any preassigned $\epsilon > 0$ there exists an integer n and a critical region in the product-space of n independent observations such that the probabilities of the errors of the first and second kind corresponding to this region are simultaneously smaller than ϵ. A sufficient condition which for a certain type of consistent test is also necessary is established. The condition is satisfied whenever the disjoint sets M and M' are closed and compact with respect to a certain suitable topology introduced on $\mathfrak{M}(\mathfrak{B})$. Thus for instance H_0 can be consistently tested against H_1 if M and M' contain only a finite number of measures or if the measures in M resp. M' depend continuously on a parameter ranging over a closed and bounded subset of some Euclidean space.

3. Effect of Linear Truncation in a Multinormal Population.

Z. William Birnbaum, University of Washington, Seattle.

Let $(X, Y_1, Y_2, \cdots, Y_{n-1})$ have a non-singular n-dimensional normal probability density $f(X, Y_1, Y_2, \cdots, Y_{n-1})$ for which all parameters are given, and let $\varphi(X, Y_1, Y_2, \cdots, Y_{n-1})$ be the probability density obtained from f by truncation along a given hyperplane $\varphi = Cf$ for $a_1Y_1 + \cdots + a_{n-1}Y_{n-1} \leqq aX + b$, $\varphi = 0$ elsewhere. What is the marginal distribution of X for this truncated distribution? This question can be answered by using a set of tables with only two parameters. These tables make it also possible to solve problems such as: determine the plane of truncation so that the marginal distribution of X has certain required properties. (This paper was prepared under the sponsorship of the Office of Naval Research.)

4. Statistical Problems in the Theory of Counters. (Preliminary Report).

Colin R. Blyth, University of California, Berkeley.

The assumptions made about counter action and distribution of incident particles are the same as those of B. V. Gnedenko [On the theory of Geiger-Müller counters, *Journ. Exper i Teor. Phiz*, Vol 11 (1941)]. The distribution of the number X of particles registered during a given time $(0, t)$ is found explicitly, in terms of the density $a(v)$ of incident particles at time v. The problem considered is that of estimating the parameters of $a(v)$. For the special case $a(v) = a$, the distribution of X reduces to $P\{X = x\} = a^x(t - x\tau)^x \exp$

$\{-a(t-x\tau)\}/x! + \exp\{-a(t-x\tau)\}\Sigma_{i=0}^{x-1} a^i[t-x\tau]^i/i! - \exp\{-a[t-(x-1)\tau]\}\Sigma_{i=0}^{x-1} a^i[t-(x-1)\tau]^i/i!$ for $x = 1, 2, \cdots, s = \left[\frac{t}{\tau}\right]$; $P\{X=0\} = e^{-at}$; $P\{X=s+1\} = 1 - \exp\{-a(t-s\tau)\}\sum_{i=0}^{s} a^i[t-s\tau]^i/i!$; $P\{X > s+1\} = 0$ This distribution has been found in another problem by J. Neyman [*On the problem of estimating the number of schools of fish*, submitted to Statistical Series, Univ of Calif press]. For this special case the maximum likelihood estimate $\hat{a}$ of a is found to be given by $\hat{a}\tau \exp(\hat{a}\tau) = \{1 + \tau/(t - x\tau)\}^x x\tau/(t - x\tau)$. If $\tau/(t - x\tau)$ is small, as will usually be the case, $\hat{a}$ will be close to the estimate $x/(t - x\tau)$ usually used for a.

5 **Some Two-Sample Tests.** Douglas G Chapman, University of California, Berkeley

Let X, Y be random variables normally distributed with means ξ, η, variances σ_1, σ_2 respectively. The two sample procedure formulated by Stein to obtain a test with power independent of σ, for the hypothesis $\eta = \xi_0$ is used here to determine a test for the hypothesis $\frac{\xi}{\eta} = r$ (r any pre-assigned real number). The size and power of this test are independent of σ_1 and σ_2. The two sample procedure may be extended to the more general case of testing the hypothesis of equality of means of several normal populations, the variances being unknown. Approximate tests are obtained for this case Finally it is shown that this two sample procedure can be used to select that normal population, of several, with the greatest mean the rule of selection having a preassigned level of accuracy (This paper was prepared with the partial support of the Office of Naval Research)

6. **Minimum Variance in Non-Regular Estimation.** R. C. Davis, U. S. Naval Ordnance Test Station, Inyokern.

The Cramér-Rao inequality for the minimum variance of a regular estimate of an unknown parameter of a probability distribution is extended to a broad class of non-regular types of estimation. The theory is developed only for the case in which a probability density function and a sufficient statistic for the unknown parameter exist For every non-regular estimation problem included in the above class, it is proved that there exists a unique unbiased estimate which attains minimum variance, and a method is given for obtaining the sample estimate. Examples are given; such as, the rectangular distribution, a class of truncated distributions, etc.

7 **Auxiliary Random Variables.** Mark W. Eudey, California Municipal Statistics, Inc., San Francisco

In testing hypotheses concerning discontinuous random variables it is not possible to find regions of arbitrary size, and so if we compare two critical regions, selection between them on the basis of the usual criteria of the Neyman–Pearson theory of testing hypotheses may be confused by the difference in their sizes. This difficulty may be avoided by allowing the statistician to use a mixed strategy in such cases, and make his decision to accept or reject the hypothesis depend upon an independent auxiliary random variable. For example, if K is a binomial variable, and U has a uniform distribution $(0, 1)$, then $Z = K + U$ may be used to test hypotheses concerning the binomial parameter, and regions of any size may be found. For the binomial case this procedure leads to a class of uniformly most powerful tests for one-sided alternatives, and to uniformly most powerful unbiased tests for two-

sided alternatives. Similar results are obtained for other common discontinuous variables, and the same device may be used in considering confidence regions and decision functions for such variables. (This paper was prepared with the partial support of the Office of Naval Research.)

8. **Estimation in Truncated Samples.** MAX HALPERIN, The Rand Corporation, Santa Monica, California.

A death process is considered which starts with n individuals of zero age, each following the mortality law, $f(x, \theta)$. That is,

$$F(t) = Pr\{\text{Age at death} < t\} = \int_0^t f(x, \theta)\, dx,$$

where $f(x, \theta)$ is a probability density. We suppose we truncate the process at a fixed time, T, and wish to estimate θ when

a) individuals who die are not replaced, and

b) individuals who die are replaced by individuals of zero age following the mortality law, $f(x, \theta)$.

In both cases, it is found that, under mild conditions, estimation by Maximum Likelihood gives optimum estimates. The estimates are best in the sense of being asymptotically normally distributed and of minimum variance for large samples.

The proofs are given for the case of a single parameter, but can be extended to the multiparameter case. Examples are given.

9. **Some Problems in Point Estimation.** J. L. HODGES, JR. AND E. L. LEHMANN, University of California, Berkeley.

Some point estimation problems are considered in the light of Wald's general theory. It is shown that when the loss function is convex, one may restrict consideration to nonrandomized estimates based on sufficient statistics. Minimax estimates are obtained in a number of cases connected with the binomial and hypergeometric distributions, and with some non-parametric problems. Some prediction problems are also considered. (This paper was prepared with the partial support of the Office of Naval Research.)

10. **Completeness in the Sequential Case.** E. L. LEHMANN AND C. STEIN, University of California, Berkeley.

Recently, in a series of papers, Girshick, Mosteller, Savage and Wolfowitz have considered the uniqueness of unbiased estimates depending only on an appropriate sufficient statistic for sequential sampling schemes of binomial variables. A complete solution was obtained under the restriction to bounded estimates. This work, which has immediate consequences with respect to the existence of unbiased estimates with uniformly minimum variance, is extended here in two directions. A general necessary condition for uniqueness is found, and this is applied to obtain a complete solution of the uniqueness problem when the random variables have a Poisson or rectangular distribution. Necessary and sufficient conditions are also found in the binomial case without the restriction to bounded estimates. This permits the statement of a somewhat stronger optimum property for the estimates, and is applicable to the estimation of unbounded functions of the unknown probability.

11. **The Ratio of Ranges.** RICHARD F. LINK, University of Oregon, Eugene.

The distribution of the ratio of two ranges from independent samples drawn from a normal population is given analytically for n_1 and $n_2 \leqq 3$. A table of percentage values, R,

is given for $\alpha = .005, .01, .025, .05, .10$ and for all combinations of n_1 and n_2 up to 10, where $\alpha = \Pr(w_1/w_2 > R)$ and w_1 and w_2 are the observed ranges. (This paper was prepared under the sponsorship of the Office of Naval Research.)

12. Some Problems Arising in Plant Selection and the Use of Analysis of Variance. STANLEY W. NASH, University of California, Berkeley.

The yields of many (m) varieties are compared in a field trial. A few varieties having the highest and lowest yields in this trial are selected for further testing. What chance is there that the first trial will give a significant result, the second trial not? Let ξ_i denote the true mean yield of the ith variety, and assume that the ξ_i are themselves normally, independently distributed with variance σ_1^2. Let P_k $(k = 1, 2)$ denote the probability of a significant result in the kth trial, using the F-test. For fixed $\sigma_1^2 > 0$, $\lim_{m\to\infty} P_1 = 1$. (See Nash, *Annals of Math. Stat.*, Vol. 19 (1948), p. 434.) Now let $\sigma_1^2 > 0$ take on a decreasing sequence of values as m increases. If $\frac{1}{\sigma_1^2 g(m)} = 0\left(\frac{E(F)}{\sigma_F}\right)$, then $\lim_{m\to\infty} P_1 = 1$. Here $1 + \sigma_1^2 g(m) = \frac{E(\text{numerator of } F)}{\sigma_0^2\ (= \text{error variance})}$. Also $\lim_{m\to\infty} P_2 < 1$ if and only if $\sigma_1^2 = 0\left(\frac{1}{\sqrt{\log m}}\right)$. For $\sigma_1^2 = 0\left(\frac{1}{\sqrt{\log m}}\right)$, $\lim_{m\to\infty} P_2 = \alpha$, the level of significance used. Thus, corresponding to any m, however large, one can find values of σ_1^2 for which the chances are considerable (or even approaching $1 - \alpha$), that the two field trials will give opposite conclusions when the F-test is used.

13. Asymptotic Properties of the Wald-Wolfowitz Test of Randomness. GOTTFRIED E. NOETHER, Columbia University, New York.

Let $a_1, \cdots, a_n$ be observations on the chance variables $X_1, \cdots, X_n$. Wald and Wolfowitz (*Annals of Math. Stat.*, Vol. 14 (1943), pp. 378–388) have shown how the statistic $R_h = \Sigma_{i=1}^{n} x_i x_{i+h}$, $(x_{n+j} = x_j)$, can be used to test the null hypothesis that the X_i, $(i = 1, \cdots, n)$, are independently and identically distributed by considering the distribution of R_h in the subpopulation of all permutations of the a_i. In the present paper it is shown that when the null hypothesis is true this distribution of R_h is asymptotically normal provided $\Sigma_{i=1}^{n} (a_i - \bar{a})^r / [\Sigma_{i=1}^{n} (a_i - \bar{a})^2]^{r/2} = 0[n^{(2-r)/4}]$, $(r = 3, 4, \cdots)$, a condition which is satisfied with probability 1 if the a_i are independent observations on the same chance variable X having positive variance and a finite absolute moment of order $4 + \delta$, $(\delta > 0)$. Conditions are given for the consistency of the test based on R_h when under the alternative hypothesis observations are drawn independently from changing populations. In particular a downward trend and a regular cyclical movement are considered, both for ranks and original observations. For the special case of a regular cyclical movement of known length the asymptotic relative efficiency of the rank test with respect to the test performed on original observations is found. It is shown that when using ranks, R_h is asymptotically normal under the alternative hypothesis provided $\liminf_{n\to\infty} \operatorname{var}(n^{-5/2} R_h) > 0$. This asymptotic normality of R_h is used to compare the asymptotic power of the R_h-test with that of the Mann T-test (*Econometrica*, Vol. 13 (1945), pp. 245–259) for the case of a downward trend.

14. On the Similar Regions of a Class of Distributions. STEFAN PETERS, University of California, Berkeley.

The class of distributions considered is essentially the class of those distributions of n variables which, by a suitable transformation of the variables and the parameter, can be transformed into distributions defined in the whole R_n for which the parameter is a location

parameter. These regions satisfy a certain partial differential equation. The transformed distributions of the variables $y_1, \cdots y_n$ and parameter τ possess a class D_1 of similar regions with respect to τ which can be defined as the smallest additive class of regions which includes all regions defined by

$$g[(y_1 - y_n), \cdots, (y_n + y_n)] \geqq C$$

where g is a continuous function. The class D_1 does not exhaust all similar regions. There exists among the regions of class D_1 one which is most powerful for testing a given additional parameter σ If there exists among all similar regions a most powerful region for testing σ, then that region will be the most powerful region of class D_1

15. **Some Problems in Sequential Analysis.** CHARLES M. STEIN, University of California, Berkeley.

Wald's fundamental identity for cumulative sums is extended to dependent random variables The first derivative of this at the origin is equivalent to a result of Wolfowitz (*Annals of Math. Stat.*, Vol. 18 (1947), p. 228, Th. 7.4). Higher derivatives of this at the origin can also be obtained from linear combinations of Wolfowitz's result applied to suitable products of the original random variables. These equations yield approximate *OC* and *ASN* curves for probability-ratio tests for a simple hypothesis against a single alternative concerning some of the more usual stationary Markoff chains Bounds for the amount by which the *ASN* exceeds that of the most efficient test are also obtained. The results are applied in particular to random variables taking on only the values 0, 1 with conditional probabilities depending only on a finite number of the preceding observations. The case of linear dependence of normal random variables with fixed conditional variance is also considered.

16. **Some Aspects of Links Between Prediction Problems and Problems of Statistical Estimation.** ERLING SVERDRUP, University of Oslo.

A prediction is not taken as a probability statement about additional observations of the random variable already observed. It is presumed that the statistical interpretation of the sample will result in some action influencing the random variable subject to prediction. The probability distribution of this random variable is given for each of an a priori class of probability functions for the observed random variable and for each of a class of possible actions. "Utility" as a function of the random variable to be predicted and of the action is defined. It is shown that the problem of which action to take in order to maximize expected utility is identical with a problem of statistical inference with a uniquely defined weight function in the Wald sense. It is further shown that this procedure is adaptable to stochastic processes of a general type and this provides a means of connecting the theory of stochastic processes with the theory of statistical inference. Some examples are given to illustrate the general theory.

17. **Some Large Sample Tests for the Median.** JOHN E. WALSH, The Rand Corporation, Santa Monica, California.

Consider a large number of independent observations from continuous populations with a common median. Some non-parametric large sample tests for the population median are presented which are based on either two or three order statistics of the sample If all the populations are symmetrical, these tests are equal-tailed with specified significance level α. If the observations are a sample from a normal population, these tests have high power efficiencies. Some tests based on three order statistics are developed which also have signifi-

cance level α if all the populations are not symmetrical; however, in this case the resulting test is one-tailed instead of equal-tailed. Using these tests for situations where the populations are believed to be symmetrical furnishes a safety factor with respect to Type I error. Tests are presented for the special case where each population is either symmetrical or skewed in a specified direction. If the populations are not symmetrical the significance level distribution is $.4\alpha$ to one tail and $.6\alpha$ to the other, rather than $.5\alpha$ to each tail. Also some non-parametric large sample tests of whether a sample is from a symmetrical population are derived. These tests are based on three order statistics of the sample and have bounded significance levels.

18. Continuous Sampling Plans from the Risk Point of View. Zivia S. Wurtele, Stanford University, California.

The quality of a lot can be improved by a screening process whereby the defective items found during inspection are replaced by non-defective items The type of sampling plan adopted will generally depend upon the cost of inspecting items, the number of defective items in the lot prior to inspection, and the loss due to defective items remaining in the lot after inspection. The loss if the lot is accepted after d defectives are found in a sample of n items is equal to $c(n) + h(D)$ where D is the number of defectives left in the lot and $c(n)$ is the cost of inspecting n items. An inspection procedure S is defined by a set of stopping points $\{(d, n)\}$. Let $r(p, S)$ be the expected loss if p is the probability of a defective and the procedure S is used It is assumed that the lot is obtained from a binomial population For any a priori distribution $F(p)$, a Bayes procedure is one which minimizes the expected risk,

$$\int_0^1 r(p,S)dF(p).$$

A systematic method of obtaining Bayes solutions exists, but the computations are formidable. Under fairly general conditions the Bayes solutions are shown to be multiple sampling plans, in which the size of the ith sample depends upon the number of defectives in the $(i-1)$st sample. In particular, if the production is in a state of statistical control, a Bayes solution is a fixed sample size It is also shown that for most reasonable loss functions, there exists no mini-max procedure which is uniformly better than the trivial one, namely, the Bayes procedure if $p = 1$.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Irving Burr has been promoted to a full professorship at Purdue University.

Dr. D. A. S. Fraser, who received his Ph.D. degree at Princeton University in June, has accepted a position as Instructor of Mathematics at the University of Toronto.

Dr. H. K. Hartline, formerly at the Johnsen Research Foundation of the University of Pennsylvania, has accepted an appointment as chairman of the Thomas Jenkins Department of Biophysics, Johns Hopkins University.

Dr. Leo Katz has been promoted to an associate professorship in the Mathematics Department of Michigan State College, East Lansing, Michigan.

Professor D. D. Kosambi of Tata Institute for Fundamental Research, Bombay, India served as Visiting Professor at the University of Chicago for the Winter Quarter.

Dr. H. G. Landau has resigned his position with the Ballistic Research Laboratories and is now a Research Associate with the Committee on Mathematical Biology, at the University of Chicago.

Mr. Allen L. Mayerson, formerly an Associate in the Division of Statistics and Research of the Institute of Life Insurance at New York City, has accepted a position with the National Surety Corporation of New York.

Mr. Raymond P. Peterson, who has been an Assistant in the Mathematics Department of the University of California at Los Angeles and also a graduate student there, has accepted a position with the Institute for Numerical Analysis at Los Angeles.

Professor Edwin J. G. Pitman has returned to the Mathematics Department of the University of Tasmania after spending about a year and a half in the United States. From February to June of 1948 he was at Columbia University as visiting Professor of Mathematical Statistics. The rest of the time was spent at North Carolina and Princeton.

A. Ananthapadmanabha Rau has returned to India after studying at the Statistical Laboratory in Ames, Iowa. In addition to heading the Department of Statistics and Agriculture Meterology of the Government of the State of Mysore, India, he is working on sampling design of experiments, and climatology and teaching statistics and climatology at the College of Agriculture.

Dr. Andrew Sobczyk of Watson Laboratories has been appointed to an assistant professorship at Boston University.

Assistant Professor S. L. Thompson of Alabama Polytechnic Institute has been promoted to an associate professorship.

William J. Youden is acting as Assistant Chief of the Statistical Engineering Section of the National Bureau of Standards and as special advisor to the Director on the problems of statistical and mathematical design of major experiments in physics, chemistry and engineering.

Two Doctorates in Mathematical Statistics were awarded at the University of North Carolina in June, 1949. The recipients were Uttam Chand, who has now been appointed Assistant Professor of Mathematics at Boston University, and Ralph A. Bradley, who will be Assistant Professor of Mathematics at McGill University.

The Educational Testing Service, Princeton, N. J., announces the appointment of Elbert Lee Hoffman and William Edward Kline as ETS Psychometric Fellows for 1949–50 for graduate study in psychology at Princeton University. Mr. Hoffman is a graduate of the University of Oklahoma, and Mr. Kline has received both his bachelor's and master's degree from Yale University. Bert F. Green, Jr. and Warren S. Torgerson have received reappointments as ETS Psychometric Fellows. Each Fellow carries a full program of graduate study in psychology at Princeton University, including basic work in experimental and theoretical psychology. Special training is also given in mathematical statistics and modern quantitative methods as applied to psychological problems in such fields as learning, testing and attitude measurement, as well as in the techniques of developing aptitude and achievement tests. In addition to the graduate program in psychology, each Fellow spends part-time in training and research work with the Educational Testing Service.

Preliminary Actuarial Examinations

Prize Awards

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of Part 2 of the 1949 Preliminary Actuarial Examinations are as follows:

First Prize of $200

Moran, Joseph W	Yale University

Additional Prizes of $100

Farmer, Thurston P., Jr.	State University of Iowa
Haakenstad, Dale L	University of Michigan
Hauke, William V	University of Michigan
Lordan, Joseph D.	Massachusetts Institute of Technology
Mayberry, John P.	University of Toronto
Murch, Alan D	University of Toronto
White, William A	Dartmouth College
Zemach, Ariel.	Harvard University

The Society of Actuaries has authorized a similar set of nine prize awards for the 1950 Examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1. *Language Aptitude Examination.*
(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning)

Part 2. *General Mathematics Examination.*
(Algebra, trigonometry, coordinate geometry, differential and integral calculus.)

Part 3. *Special Mathematics Examination.*
(Finite differences, probability and statistics.)

The 1950 Preliminary Actuarial Examinations will be administered by the Educational Testing Service at centers throughout the United States and Canada on May 19, 1950. The closing date for applications is March 15, 1950.

Detailed information concerning the Examinations can be obtained from:

The Society of Actuaries
208 South LaSalle Street
Chicago 4, Illinois

New Members

The following persons have been elected to membership in the Institute

(March 1, 1949 to May 31, 1949)

Alcantara de Oliveira, Eduardo, Ph.D., (Univ de Sao Paulo) Professor, Faculdade de Filosofia, University of Sao Paulo, *Rua Sergipe, 90-Ap. 82, Sao Paulo, Brazil.*

Ashby, Wallace L., A.B. (George Washington Univ.) Agricultural Statistician, *3746 Jocelyn Street, Washington 15, D. C.*

Balley, Edward W., B.Ch. (Ohio State Univ.) Quality Control Supervisor, Carbide and Carbon Chemicals Corporation, Y-12 Plant, *101 Maylan Lane, Oak Ridge, Tennessee.*

Berger, Agnes P., Ph.D. (Budapest) *10 Park Avenue, New York, New York.*

Brown, Walter C., B.S. (Colorado A&M College) Graduate Assistant, Department of Mathematics, University of Oklahoma, *1180 Trout, Norman, Oklahoma.*

Calvin, Lyle D., B.S. (Univ. of Chicago) Research Graduate Assistant, Institute of Statistics, North Carolina State College, Raleigh, North Carolina

Carlyle, Charles G., B.S. (Univ. of Illinois) Graduate student at University of Illinois, *C-32 Stadium Terrace, Champaign, Illinois.*

Chen, Yu-nien, M.A. (Harvard) Graduate student, Harvard University, *147-60, Apt. D, Charter Road, Jamaica 2, New York.*

Clark, Fred J., Jr., B.S. (Colorado A&M College) Graduate Assistant at University of Illinois, Department of Mathematics, *61 A Court G, Stadium Terrace, Champaign, Illinois.*

Cohen, Samuel E., M.A. (Univ. of Pennsylvania) Statistician, U. S. Bureau of Labor Statistics, *49 Galveston St., S.W., Washington 20, D. C.*

Cole, Randal H., Ph.D. (Univ. of Wisconsin) Associate Professor, University of Western Ontario, London, Canada.

Comrey, Andrew L., Ph.D. (Univ. of Southern Calif.) Assistant Professor of Psychology, University of Illinois, Urbana, Illinois.

Cook, Ellsworth B., B.S. (Springfield College) Head of Visual Screening Devices Research and Statistics Facility, U. S. Naval Medical Research Laboratory, *Box 45, Submarine Base, New London, Connecticut.*

Cox, David R., Ph.D. (Leeds, England) Statistician, Wool Industries Research Association, *2 Sunset Avenue, Leeds 6, Yorks, England.*

Denbow, Carl H., Ph.D. (Univ. of Chicago) Associate Professor of Mathematics, U. S. Naval Postgraduate School, Annapolis, Maryland.

Dillon, Gregory M., A.B. (Long Island Univ.) Statistician, Pension Statistics Section,

Treasury Department, E I DuPont de Nemours & Co , *1331 Cedar Street, Wilmington, Delaware.*

Duarte, Geraldo Garcia, Licenciado em Matematica (Faculdade de Filosofia de S Bento) Assistente da Faculdate de Higiene e Saude Publica, Caixa Postal 99B, Sao Paulo, Brazil.

Dudman, John A., B A. (Reed College) Graduate student, Columbia University, *56 West 70th St., New York 23, New York.*

Edelson, Howard, B.A (Ohio State Univ , Columbus, Ohio) Graduate student and Graduate Assistant, Ohio State University, *794 S 18th St , Columbus 6, Ohio.*

Feron, R., Licencie es Sciences, (Univ of Paris) Attache de Recherce, *13 rue des Feuillantines, Paris V, France*

Franck, Edward Michel, Inpineur A.I A., Professor of the Royal Military School, *104 Rue Pere Devroye, Woluwe St. Pierre, Belgium.*

Garritsen, Florence M., B A (Univ of Michigan) Research Assistant, General Motors Corp., *5151 Lillibridge Ave , Detroit 13, Michigan.*

Gelsomini, Thea, Ph D (Univ. of Bocconi, Milano) Assistician of Statistics at Department of Statistics, University of Bocconi, *Via A Stoppi, N 10, Milano, Italy.*

Goudswaazd, G, Ph D (Univ of Leiden) Director, Permanent Office, International Statistical Institute and Lecturer of Statistics, Rotterdam School of Economics and Free University of Amsterdam, 2 Oostduinlaan, The Hague, Netherlands.

Gucker, Frank Fulton, A B (Harvard Univ) Statistical Engineer, Remington Arms Co , Inc , *3175 Main Street, Bridgeport 6, Connecticut.*

Haberman, Sol, B.A. (Brooklyn College) Assistant Visiting Professor of Sociology, University of Puerto Rico, *187 Avenida los Flamboyanes, Rio Piedras, Puerto Rico*

Heimbach, Ernest E., M.B A (New York Univ) Professor of Economics, Bergen College, Teaneck, New Jersey, *55 West 11th Street, New York 11, N. Y*

Ishii, Shigeru, B.A. (Univ of Ill) Student at University of Illinois, *320-1 Peabody Drive, Parade Ground Units, Champaign, Illinois*

Jackson, James Edward, M.A (Univ of N C) Statistician, Color Control Dept , Eastman Kodak Company, *200 Pershing Drive, Rochester, New York.*

Jaspen, Nathan, Ph D (Pennsylvania State College) Research Assistant, Department of Psychology, Pennsylvania State College, State College, Pennsylvania

Jonhagen, Sven, Fil Lic. (Univ. of Stockholm) Chief Actuary and Assistant Teacher in Statistics at the University of Stockholm, *Tegnergatan 36, Stockholm, Sweden*

Kiefer, Jack C., M.S (Mass Inst. of Tech.) Student, Department of Mathematical Statistics, Columbia University, *3826 Middleton Avenue, Cincinnati 20, Ohio.*

Kraft, Charles Hall, B A. (Mich. State College) Instructor, Mathematics Department, Michigan State College, *707D Chestnut Road, East Lansing, Mich*

Mayne, John W., M Sc (Brown Univ.) Graduate student in Mathematical Statistics, 330 Furnald Hall, Columbia University, New York 27, New York.

McCabe, William J., M.A. (George Washington Univ.) Chief Statistician, Transportation Corps, Department of the Army, *1725 South Oakland Street, Arlington, Virginia.*

Medin, Knut H., M A (Univ of Uppsala) Assistant, Statistical Institute, University of Uppsala, Odinslund 2, Uppsala, Sweden.

Mewborn, A. Boyd, Ph.D. (Calif Inst of Tech.) Associate Professor of Mathematics and Mechanics, *P.O. Box 1748, Monterey, California.*

Minton, Paul D., M S (Southern Methodist Univ.) Graduate student, University of North Carolina, P O Box 634, Chapel Hill, North Carolina

Morris, Doris N., M A. (Columbia Teachers College) Economics Assistant, Western Electric Co , *101 West 72nd St., New York 23, New York*

Morris, Robert H., B A (Swarthmore) Development Engineer, Color Control Department, Eastman Kodak Co , Rochester, New York.

Rajalakshman, D. V., M.Sc. (Madras Univ.) Head, Department of Statistics, University of Madras, Madras 5, S. India.

Rudy, Norman, M.B.A. (Univ. of Chicago) Scientist, Ordnance Research Project, University of Chicago and Instructor in Economics, Roosevelt College, *7105 S. Crandon Avenue, Chicago 49, Illinois.*

Sakk, Kaarel, Fil. kand. (Univ. of Stockholm) Officer at Research Bureau of the State Foodstuffs Commission, Ostermalmsgatan 87 o.g. III, Stockholm, Sweden.

Singh, Jagjit, B.A. (Punjab Univ.) Superintendent Transportation, E. I. Railway, Dinapore, *c/o B.S. Bugga Esqr., Post Office Box 441, Calcutta, India.*

Starr, Henry H., Ph.D. (Univ. of Vienna) Research Manager, Converted Rice, Inc., P.O. Box 1752, Houston 1, Texas.

Sverdrup, Erling, Actuarian (Univ. of Oslo) Lecturer in Mathematical Statistics, Institute of Mathematics, University of Oslo, Oslo, Norway.

Talacko, Joseph V., Ph.D. (Charles Univ., Prague) Assistant Professor of Mathematics, Marquette University, *2503 So. 10th Street, Milwaukee 7, Wisconsin.*

Templeton, James G. C., A.M. (Princeton Univ.) Graduate student at Princeton University, Fine Hall, Princeton University, Princeton, New Jersey.

Vaughan, Elizabeth, B S. (Univ. of Washington) Statistician, U.S. Fish and Wildlife Service, 2725 Montlake Boulevard, Seattle 2, Washington.

Wilkinson, Bryan, M.A. (Univ. of Nebraska) Personnel Research Specialist, Prudential Insurance Co., Western Home Office, *4130-B Muirfield Road, Los Angeles 43, California.*

Yevick, Mariam A. L., Ph.D. (Mass. Inst. of Tech.) Staff, Division of Statistical Engineering, National Broadcasting System, *921 Hudson St., Hoboken, New Jersey.*

REPORT ON THE BERKELEY MEETING OF THE INSTITUTE

The thirty-ninth meeting and fifth regional West Coast meeting of the Institute of Mathematical Statistics was held on the Berkeley campus of the University of California, from Thursday June 16 through Saturday June 18, 1949. The session on June 17 was held jointly with the Biometrics Section of the American Statistical Association and the Biometric Society (Western N. A. Region). Sixty-six persons registered, including the following fifty members of the Institute:

Jane F. Andrian, G A Baker, Z. Wm. Birnbaum, Colin R Blyth, Albert H Bowker, Paul T. Bruyere, Chin Long Chiang, Edwin L Crow, John H Curtiss, R C. Davis, Carl H. Denbow, W J Dixon, Mary Elveback, Mark Eudey, Edward A. Fay, Evelyn Fix, William R Gaffey, II H Germond, M. A Girshick, Jack Gysbers, Max Halperin, J. L. Hodges, Jr , John M Howell, Harry M Hughes, Cuthbert Hurd, Terry A. Jeeves, Mark Kac, H. S Konijn, George M Kuznets, Erich L Lehmann, Richard F Link, Michel Loève, Frank Massey, Lincoln E Moses, Edith Mourier, Stanley W Nash, J. Neyman, Edward Paulson, Stefan Peters, Raymond P Peterson, Robert I. Piper, Gladys Rappaport, Mina Rees, David Rubinstein, Elizabeth L. Scott, Esther Seiden, Charles M. Stein, John E. Walsh, John Wishart, Zivia S Wurtele.

Those attending were welcomed at the Thursday morning session by Edward W. Strong, Associate Dean of the College of Letters and Science, University of California. Professor Z. William Birnbaum of the University of Washington presided.

The program was as follows:

1. *Recent advances in the theory of the Wishart distribution.* (Invited paper.) John Wishart, Cambridge University.
2. *Bayes, minimax, and other approaches to the multiple classification problem.* (Invited paper.) M A. Girshick, Stanford University.
3. *Some problems in sequential analysis.* Charles M. Stein, University of California, Berkeley.

Professor Jerzy Neyman of the University of California, Berkeley, presided at the Thursday afternoon session. Midway in the program there was an intermission for a tea given by the Statistical Laboratory, University of California. The program was as follows:

1. *Completeness in the sequential case.* E. L. Lehmann and C. M. Stein, University of California, Berkeley
2. *Some large sample tests for the median* John E Walsh, The Rand Corporation.
3. *Continuous sampling plans from the risk point of view* Zivia S Wurtele, Stanford University
4. *Some problems in point estimation* J L Hodges, Jr and E L Lehmann, University of California, Berkeley
5. *Minimum variance in non-regular estimation* R. C. Davis, U. S Naval Ordnance Test Station, Inyokern.
6. *Some aspects of links between prediction problems and problems of statistical estimation* Erling Sverdrup, University of Oslo.

7. *Extension of a theorem of Blackwell.* (By title). Edward W Barankin, University of California, Berkeley
8. *Some two-sample tests.* (By title). Douglas G. Chapman, University of California, Berkeley.
9. *On the existence of consistent tests.* (By title). Agnes Berger, Columbia University.

Professor F. W. Weymouth of Stanford University presided at the Friday morning session on biometrics. The program was as follows:

1. *Statistical problems arising from research in tuberculosis.* Martha and Paul T. Bruyere, U. S Public Health Service.
2. *Correlation of variability with growth rate in fish and mollusks.* F W. Weymouth, Stanford University.
3. *Some problems arising in plant selection and the use of analysis of variance* Stanley W Nash, University of California, Berkeley.
4. *Studies of resistance of strawberry varieties and selections to verticillum wilt* R. E Baker and G. A. Baker, University of California, Davis.
5. *A uniformity trial on unirrigated barley of ten years duration with implications for field trial designs.* F. J. Veihmeyer, M. R. Huberty, and G. A. Baker, University of California, Davis and Los Angeles.

On Friday afternoon those attending the meeting were entertained at a picnic luncheon at Stanford University, given by the Department of Statistics, Stanford University.

Professor C. B. Morrey, Jr., of the University of California, Berkeley, presided at the Saturday morning session. The program consisted of the following invited papers:

1. *Methods for getting limiting distributions.* Mark Kac, Cornell University.
2. *Almost certain convergence.* Michel Loève, University of California, Berkeley

At 11 o'clock Saturday morning a business session was held, under the chairmanship of Professor Jerzy Neyman of the University of California, Berkeley, for the purpose of discussing future West Coast meetings. Plans for reviving the Statistical Research Memoirs were also discussed.

On Saturday afternoon a final session for contributed papers was held under the chairmanship of Professor Albert H. Bowker of Stanford University. The program was as follows:

1. *Effect of linear truncation in a multinormal population.* Z. William Birnbaum, University of Washington
2. *Estimation in truncated samples.* Max Halperin, The Rand Corporation.
3. *On the similar regions of a class of distributions.* Stefan Peters, University of California, Berkeley
4. *Auxiliary random variables.* Mark W. Eudey, California Municipal Statistics.
5. *The ratio of ranges* Richard F. Link, University of Oregon.
6. *Statistical problems in the theory of Geiger counters.* Colin R. Blyth, University of California, Berkeley.
7. *Asymptotic properties of the Wald-Wolfowitz test of randomness.* (By title). Gottfried E. Noether, Columbia University

J. L. HODGES, JR.
Assistant Secretary

LOCALLY BEST UNBIASED ESTIMATES[1]

By E. W. Barankin

University of California, Berkeley

Summary. The problem of unbiased estimation, restricted only by the postulate of section 2, is considered here. For a chosen number $s > 1$, an unbiased estimate of a function g on the parameter space, is said to be best at the parameter point θ_0 if its sth absolute central moment at θ_0 is finite and not greater than that for any other unbiased estimate. A necessary and sufficient condition is obtained for the existence of an unbiased estimate of g. When one exists, the best one is unique. A necessary and sufficient condition is given for the existence of only one unbiased estimate with finite sth absolute central moment. The sth absolute central moment at θ_0 of the best unbiased estimate (if it exists) is given explicitly in terms of only the function g and the probability densities. It is, to be more precise, specified as the l.u.b. of certain set $\mathfrak{A}$ of numbers. The best estimate is then constructed (as a limit of a sequence of functions) with the use of only the data (relating to g and the densities) associated with any particular sequence in $\mathfrak{A}$ which converges to the l.u.b. of $\mathfrak{A}$.

The case $s = \infty$ is considered apart. The case $s = 2$ is studied in greater detail. Previous results of several authors are discussed in the light of the present theory. Generalizations of some of these results are deduced. Some examples are given to illustrate the applications of the theory.

1. Introduction. Let Ω be a space of points x, and μ be a totally additive measure defined on a σ-field $\mathcal{F}$ of subsets of Ω. Let $\mathfrak{P} = \{p_\theta, \theta \in \Theta\}$ be a family of probability densities in Ω with respect to the measure μ. Θ is any index set; we lay down no conditions on its structure. We are concerned here with the existence and characterization of unbiased estimates of a real-valued function g on Θ, which are in some suitable sense "best" for a prescribed parameter point θ_0. That is, a real-valued, measurable (μ) function f_0 on Ω such that

$$\int_\Omega f_0 p_\theta \, d\mu = g(\theta), \qquad \theta \in \Theta, \tag{1}$$

and which satisfies a specified criterion of bestness for $\theta = \theta_0$. This criterion is usually taken to be

$$\int_\Omega (f_0 - g(\theta_0))^2 p_{\theta_0} \, d\mu \leqq \int_\Omega (f - g(\theta_0))^2 p_{\theta_0} \, d\mu, \qquad f \in \overline{\mathfrak{M}}, \tag{2}$$

where $\overline{\mathfrak{M}}$ denotes the class of all unbiased estimates of g; i.e., the class of all f satisfying (1). The obvious advantage in the definition (2) is the algebraic

[1] This article was prepared while the author was under contract with the Office of Naval Research.

pliability. The obvious disadvantage is that $\overline{\mathfrak{M}}$ may contain no estimate with finite variance (cf. section 9).

For the investigation of the fundamental questions, posed above, relating to unbiased estimates, we shall not restrict ourselves to (2). We consider chosen and fixed, a number $s > 1$, and lay down the

DEFINITION. *$f_0 \in \overline{\mathfrak{M}}$ is best at θ_0 if*

$$\infty > \int_\Omega |f_0 - g(\theta_0)|^s p_{\theta_0}\, d\mu \leqq \int_\Omega |f - g(\theta_0)|^s p_{\theta_0}\, d\mu, \qquad f \in \overline{\mathfrak{M}}.$$

With this, and under the condition of a rather natural postulate on $\mathfrak{P}$ (cf. section 2), we exhibit a necessary and sufficient condition for the existence of an unbiased estimate of g having a finite sth absolute central moment at θ_0.[2]

Except for the discussion, in section 3, of the case in which g is constant on Θ, we do not consider directly the estimation of g, but rather that of $h = g - g(\theta_0)$. Lemma 1, of section 2, gives the solution of the problem for g when that for h is known After section 3, it is assumed exclusively that h is not $\equiv 0$, except where the contrary is explicitly stated.

In case s is finite, the existence theorem section 4, Theorem 2, asserts also the uniqueness of the best unbiased estimate of h. It is interesting to observe the similarity between the proof of this uniqueness and Fisher's proof of the (what might be called) asymptotic uniqueness of an efficient estimator [2 pp. 704, 705]. The case $s = \infty$[3] is discussed in section 5; in this case we find that, in general, the best estimate is not unique. However, for s both finite and infinite, and as well when g is constant ($\therefore h \equiv 0$), we give a necessary and sufficient condition that there be a unique unbiased estimate with finite s.a.c.m.[4] (cf. section 4, Corollary 2-1, and section 5, Theorem 3 (iii)).

Theorem 2 determines the s.a.c.m. of the best estimate as the l.u.b. of a set of numbers given explicitly; and thereby, in particular, throws open the class of all lower bounds of the minimum s.a.c.m. Investigations after such lower bounds, in the classical case $s = 2$, have led to the well-known results of Cramér-Rao [3 p. 480, (32.3.3)], and Bhattacharyya [4, p. 3, (1.10)]. In section 6, which is devoted to obtaining various special lower bounds, we show how those particular bounds fall out. It should be remarked, however, that our conditions on $\mathfrak{P}$ are in general different from those of the above authors.

---

[2] For the case $s = 2$ an alternative existence condition, antedating these results, but not yet published, has been obtained by C. Stein.

[3] If we use, in the above definition, the sth root of the sth absolute central moment, instead of the latter itself, then the bestness criterion for $s = \infty$ is the limiting criterion for $s \to \infty$; viz.,

$$\infty > \operatorname*{ess\,sup}_{x \in \Omega} |f_0 - g(\theta_0)| \leqq \operatorname*{ess.\,sup}_{x \in \Omega} |f - g(\theta_0)|, \qquad f \in \overline{\mathfrak{M}},$$

where ess. sup. refers to the measure $\nu(A) = \int_A p_{\theta_0}\, d\mu$.

[4] The abbreviation s.a.c.m. will henceforth be used to indicate sth absolute central moment at θ_0.

In section 7 we give, in Theorem 7 and its corollary, a construction of the best estimate, depending only on the knowledge of the minimum s.a.c.m. The latter, as indicated in the preceding paragraph, is always known independently of any knowledge of the best estimate. We use these results to obtain explicitly (Theorems 8 and 9) the best estimates, for arbitrary s, in two cases where we assume the minimum s.a.c.m. known. These cases, when $s = 2$, give the minimum variance as determined by the equality sign in the Cramér-Rao and Bhattacharyya inequalities, respectively.

Section 8 is given to a brief discussion of the special case $s = 2$. Finally, in section 9, we present a detailed study of an example

At the suggestion of the referee we have added an appendix in which is given a brief running description of the fundamental ideas of Banach spaces that come into use here The italicized phrases are those mentioned explicitly in the course of the paper

We shall merely mention here certain points which will be elaborated further in future communications. (1) The general theory developed here pertains as well to sequential as to nonsequential estimation; one has only to make the proper identification of Ω, $\mathcal{F}$, μ, and $\mathfrak{P}$. Moreover, as applied to sequential estimation, the theory will determine the optimum stopping regions. (2) The discussion of section 5 below can be carried through with "ess. sup." referring to the measure μ, and $\mathfrak{L}_1$ being the space of functions on Ω which are integrable (μ); and for this, no restrictions whatsoever on the densities p_θ are required (cf. the postulate of section 2), since the p_θ are elements of this $\mathfrak{L}_1$ solely by virtue of their properties as probability densities. This development would, for example, be sufficient to yield the estimate of Girshick, Mosteller, and Savage [5] in the case of sequential binomial estimation. Also, this unrestricted analysis is fundamental for the problem of similar regions (a case of the bounded unbiased estimation of a constant function). (3) For any $s > 1$ it may be observed in the result of Theorem 7 below, that the best (at θ_0) estimate depends only on a sufficient statistic, this is clear from Neyman's theorem on sufficient statistics [6], since the best estimate depends only on ratios of the density functions p_θ But more than this, Blackwell's method [7] of deriving a uniformly (over the parameter set) better unbiased estimate from a given unbiased estimate can be proved to remain valid also when the measure of dispersion is the sth absolute central moment, $s > 1$ And for this, the postulate of section 2 is not required. (4) Finally, we point out that, with the proper specializations of Θ, Cramér's theorem on the ellipsoid of concentration [8], Bhattacharyya's multidimensional inequality [9], and the extensions of the Rao, Cramér, and Bhattacharyya bounds to sequential estimation—as, for example, by Blackwell and Girshick [1], Wolfowitz [10], and Seth [20]—can be drawn from Theorem 4 below.

The inspiration for the mode of analysis in the following pages, and the major part of its substance, come from F. Riesz: his book [11 Ch. III] and the article [12] (in particular sections 8–11 thereof) In strictly mathematical terminology, Theorems 2 and 3 are given in [11] for the sequence-spaces ℓ_r ; and

Theorem 2 in [12] for the spaces $\mathfrak{L}_r$ of functions on the real interval [0, 1] with Lebesgue integrable rth powers. The proofs are given there for the case of a denumerable set Θ; in [12] an indication is given of the extension to a non-denumerable Θ. Our proof of Theorems 2 and 3, however, follows that given by Banach [13, p. 74] for the case of denumerable Θ. It is based on two results, a theorem of Hahn-Banach [13, p. 55, Theorem 4], and the representation theorem (suitable for the general type of $\mathfrak{L}_r$ that we consider) for bounded linear functionals on $\mathfrak{L}_r$ [14, p. 338, Theorem 46]. The first of these, and the representation theorem for any $r > 1$, spring in fact from the same article [12, p. 475] of Riesz. In the case $r = 1$, the representation theorem is due originally to Steinhaus [15]; in the case $r = 2$, it was developed simultaneously in 1907 by Riesz [16] and Fréchet [17].

Riesz' proofs of the sufficiency of the condition in Theorem 2 proceed by constructing an explicit sequence of functions on Ω which converge strongly in $\mathfrak{L}_r$ to the (in the present statistical terminology) best estimate. Precisely, if in Theorem 7 below, we take, for each $n = 1, 2, \cdots$, the numbers $\alpha_1^n, \alpha_2^n, \cdots, \alpha_{k_n}^n$ so that the expression

$$\frac{\left|\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)\right|}{\left\|\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}\right\|_r}$$

is maximum, then the assertion of this theorem is that of Riesz. However, Theorem 7 is established here without this strict requirement on the α_i^n. The dropping of this restriction was essential for the proofs of Theorems 8 and 9. The latter two theorems are, in fact, proved with the use of Corollary 7–1, which is an even stronger result than Theorem 7. This corollary falls out of the proof of Theorem 7 immediately, in consequence of our use of Lemma 2 for that proof. The lemma, moreover, eases the proof of Theorem 7 markedly, in doing away with the need for any differentiation.

2. Preliminary considerations. We begin then by introducing the absolutely continuous (with respect to μ) measure, defined on $\mathfrak{F}$,

$$\nu(A) = \int_A p_{\theta_0}\, d\mu, \qquad A \in \mathfrak{F}.$$

A function ϕ is summable (ν) over Ω if and only if $\phi \cdot p_{\theta_0}$ is summable (μ) over Ω; and we have

$$\int_\Omega \phi\, d\nu = \int_\Omega \phi \cdot p_{\theta_0}\, d\mu,$$

(cf. [18, pp. 36–38]). Assuming that each of the ratios

$$\pi_\theta(x) = \frac{p_\theta(x)}{p_{\theta_0}(x)}, \qquad \theta \in \Theta$$

is defined almost everywhere (μ) throughout Ω, it follows that f is an unbiased estimate of g if and only if

$$\int_\Omega f\pi_\theta \, d\nu = g(\theta), \qquad \theta \in \Theta. \tag{3}$$

We define

$$h(\theta) = g(\theta) - g(\theta_0).$$

Since

$$\int_\Omega \pi_\theta \, d\nu = 1, \qquad \theta \in \Theta,$$

it is clear from (3) that f is an unbiased estimate of g if and only if $f - g(\theta_0)$ is an unbiased estimate of h. Moreover, f is best, for g, at θ_0 if and only if $f - g(\theta_0)$ is best, for h, at θ_0.

Define

$$r = \frac{s}{s-1},$$

and let $\mathfrak{L}_r$ and $\mathfrak{L}_s$ be the spaces, normed in the usual way, of real-valued functions on Ω, with summable (ν) absolute rth and sth powers, respectively. We denote the respective norms by $\| \ \|_r$ and $\| \ \|_s$; that is, if $\xi \in \mathfrak{L}_r$ and $\eta \in \mathfrak{L}_s$,

$$\| \xi \|_r = \left(\int_\Omega | \xi |^r \, d\nu \right)^{1/r},$$

and

$$\| \eta \|_s = \left(\int_\Omega | \eta |^s \, d\nu \right)^{1/s}.$$

We note that these spaces, for $s < \infty$, are weakly compact (cf. [21]). This property will be used in the proof of Theorem 7. Also, we shall make explicit use of the representation theorem for linear functionals on $\mathfrak{L}_r$ [14, p. 338, Theorem 46].

The assumptions on $\mathfrak{P}$, or on $\mathfrak{P}_0 = \{\pi_\theta, \theta \in \Theta\}$, will now be the following.

POSTULATE: *The functions π_θ are defined almost everywhere (μ) in Ω, and are elements of $\mathfrak{L}_r$.*

The foregoing considerations combine to give the following equivalence.

LEMMA 1. *$\phi_0 + g(\theta_0)$ is an unbiased estimate of g, which is best at θ_0, if and only if (i) ϕ_0 satisfies the equations*

$$\int_\Omega \phi \cdot \pi_\theta \, d\nu = h(\theta), \qquad \theta \in \Theta, \tag{4}$$

and (ii) when ϕ is any other function satisfying (4), we have

$$\| \phi \|_s \geqq \| \phi_0 \|_s ;$$

that is, if and only if ϕ_0 is an unbiased estimate of h with minimum (finite) norm in $\mathfrak{L}_s$. The s.a.c.m. of $\phi_0 + g(\theta_0)$ is precisely $\| \phi_0 \|_s$.

Starting with section 4, we shall deal directly with the estimation of h.

3. The case of constant g. Throughout the remainder (section 4 et seq.) of this article, the function h is assumed, unless the contrary is explicitly stated, to be non-constant; that is, since $h(\theta_0) = 0$, not $\equiv 0$. We can, and shall in this section, obtain the results of the desired kind for the case of a constant function g, by a brief, direct attack.

Let $g(\theta) \equiv g_0$, a constant. Then of course $h(\theta) \equiv 0$. One unbiased estimate of g is immediately obvious, viz., $f_1(x) \equiv g_0$. The s.a.c.m. of f_1 is 0.

There will exist other[5] unbiased estimates of g with finite s.a.c.m. if and only if there exist non-null unbiased estimates, in $\mathfrak{L}_s$, of $0 \equiv h$. That is, by virtue of the isomorphism between $\mathfrak{L}_s$ and the space of linear functionals on $\mathfrak{L}_r$, there will exist an unbiased estimate of g with finite s.a.c.m., distinct from f_1, if and only if there exists a non-null functional on $\mathfrak{L}_r$ which vanishes on the elements of $\mathfrak{P}_0 = \{\pi_\theta, \theta \in \Theta\}$. And a necessary and sufficient condition that such a functional exist is that $\mathfrak{P}_0$ be not a fundamental set in $\mathfrak{L}_r$ [13, p. 58, Theorem 7].

Observe finally that, in any case, f_1 is the unique unbiased estimate of g with vanishing s.a.c.m.

We collect these results in the following statement.

THEOREM 1. *If $g(\theta) \equiv g_0$, a constant, then there is a unique best unbiased estimate of g; viz., $f_1(x) \equiv g_0$. And the s.a.c.m. of f_1 is 0.*

A necessary and sufficient condition that there exist no other unbiased estimates of g having finite s.a.c.m. is that the set $\mathfrak{P}_0$ be fundamental in $\mathfrak{L}_r$.

As an illustration of the ideas of this section, consider the following example: Ω is the real interval [0, 1]; μ is Lebesgue measure; Θ is the set of non-negative integers; and

$$p_\theta(x) = (\theta + 1)x^\theta.$$

And take $\theta_0 = 0$. Then, ν is again Lebesgue measure, and $\pi_\theta = p_\theta$ for each θ. For definiteness, take $r = 2$ (the results in this case are the same for any $r \geqq 1$). It is well-known that the non-negative integer powers of x form a fundamental set in $\mathfrak{L}_2$ on a finite real interval. That is, if ξ is a function on [0, 1], such that $\int_0^1 \xi^2 \, dx < \infty$, and if $\epsilon > 0$, then there exist an integer n and coefficients b_0,

---

[5] That is, distinct from f_1 in the sense of $\mathfrak{L}_s$; or, equivalently, differing from f_1 on a set of positive (ν) measure. Whenever, in the sequel, an equation $\xi_1 = \xi_2$ appears, for two functions ξ_1 and ξ_2 in $\mathfrak{L}_r$ or $\mathfrak{L}_s$, equality almost everywhere (ν) in Ω will be understood. It is a consequence of our postulate that if two functions on Ω are equal almost everywhere (ν), they are equal almost everywhere (ν'), where ν' is anyone of the measures $\nu'(A) = \int_A p_{\theta'} \, d\mu$, $\theta' \in \Theta$

$b_1, \cdots, b_n$ such that

$$\int_0^1 \left(\xi - \sum_{i=0}^{n} b_i x^i\right)^2 dx < \epsilon.$$

Hence, in this case an unbiased estimate with finite variance at $\theta = 0$ is unique (as well for a non-constant function g as for one which is constant over Θ; cf. section 4, Corollary 2-1).

4. The main theorem for non-constant h. We shall denote by $\mathfrak{M}_s$ the class (or the set in $\mathfrak{L}_s$) of all unbiased estimates of h that belong to $\mathfrak{L}_s$.

THEOREM 2. (i) *A necessary and sufficient condition that $\mathfrak{M}_s$ be non-empty is that there exist a constant C such that for every set of n functions $\pi_{\theta_1}, \pi_{\theta_2}, \cdots, \pi_{\theta_n}$, in $\mathfrak{P}_0$, and every set of n real numbers $a_1, a_2, \cdots, a_n$, we have, for every $n = 1, 2, \cdots$,*

$$\left|\sum_{i=1}^{n} a_i h(\theta_i)\right| \leqq C \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r. \tag{5}$$

(ii) *For every $\phi \in \mathfrak{M}_s$ we have $||\phi||_s \geqq C_0$, where C_0 is the g.l.b. of the set of admissible constants C in* (5)

(iii) *If $\mathfrak{M}_s$ is non-empty there is a unique $\phi_0 \in \mathfrak{M}_s$ with $||\phi_0||_s = C_0$. Thus, ϕ_0 is the unique unbiased estimate of h which is best at θ_0.*

The non-constancy of h clearly implies $C_0 > 0$.

The necessity of condition (5) is immediate. Suppose $\phi \in \mathfrak{M}_s$, so that ϕ satisfies equations (4); then, for any $\theta_1, \theta_2, \cdots, \theta_n$, and any real numbers $a_1, a_2, \cdots, a_n$,

$$\sum_{i=1}^{n} a_i h(\theta_i) = \int_\Omega \phi \sum_{i=1}^{n} a_i \pi_{\theta_i} \cdot dv.$$

By the Hölder inequality it follows that

$$\left|\sum_{i=1}^{n} a_i h(\theta_i)\right| \leqq ||\phi||_s \cdot \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r.$$

Hence (5) is satisfied with $C = ||\phi||_s$.

Part (ii) of the theorem is hereby proved as well.

Suppose $\mathfrak{M}_s$ non-empty, and ϕ_0, ϕ_1 in $\mathfrak{M}_s$, such that $||\phi_0||_s = ||\phi_1||_s = C_0$. Then $1/2\ (\phi_0 + \phi_1) \in \mathfrak{M}_s$ and therefore

$$1/2\ ||\phi_0 + \phi_1||_s \geqq C_0.$$

But, by the Minkowski inequality,

$$1/2\ ||\phi_0 + \phi_1||_s \leqq 1/2\ (||\phi_0||_s + ||\phi_1||_s) = C_0,$$

Hence

$$||\phi_0 + \phi_1||_s = ||\phi_0||_s + ||\phi_1||_s.$$

This equality implies $\phi_1 = \alpha \phi_0$ for some positive α. But since the norms of ϕ_0 and ϕ_1 are equal (and $\neq 0$) α must be unity Thus the uniqueness of ϕ_0 is proved.

It remains now to prove, assuming (5) satisfied, the existence of ϕ_0. Consider the functional F on $\mathfrak{P}_0$ defined by

$$F(\pi_\theta) = h(\theta).$$

The Hahn-Banach theorem alluded to in section 1 (viz., [13, p. 55, Theorem 4]) has precisely (5) as a necessary and sufficient condition for the existence of a linear functional G on $\mathfrak{L}_r$ satisfying

$$\begin{aligned} &\text{(a)}\ G(\pi_\theta) = h(\theta),\ \theta \in \Theta; \\ &\text{(b)}\ \| G \| \leqq C; \end{aligned}$$

where $\| G \|$ is the norm of G, i.e.,

$$\| G \| = \underset{\xi \in \mathfrak{L}_r}{\text{l.u.b.}} \frac{| G(\xi) |}{\| \xi \|_r}.$$

In particular, taking $C = C_0$, there is a linear functional G_0 on $\mathfrak{L}_r$ with

$$\begin{aligned} &\text{(a')}\ G_0(\pi_\theta) = h(\theta),\ \theta \in \Theta \\ &\text{(b')}\ \| G_0 \| \leqq C_0. \end{aligned}$$

But, for an element $\sum_{i=1}^{n} a_i \pi_{\theta i}$ in the linear manifold $[\mathfrak{P}_0]$ spanned by the π_θ,

$$G_0 \left(\sum_i a_i \pi_{\theta_i} \right) = \sum_i a_i h(\theta_i),$$

so that

$$\| G_0 \| \geqq \underset{\xi \in [\mathfrak{P}_0]}{\text{l.u.b.}} \frac{| G_0(\xi) |}{\| \xi \|_r} = C_0.$$

Hence (b') is replaced by the precise statement

$$\text{(b'')}\ \| G_0 \| = C_0.$$

Now the representation theorem for linear functionals on $\mathfrak{L}_r$ asserts the existence of $\phi_0 \in \mathfrak{L}_s$, such that

$$G_0(\xi) = \int_\Omega \phi_0 \cdot \xi \, d\nu,$$

and

$$\| \phi_0 \|_s = \| G_0 \| = C_0.$$

This taken with (a') establishes the existence of $\phi_0 \in \mathfrak{L}_s$ satisfying

$$\begin{cases} \int_\Omega \theta_0 \pi_\theta \, d\nu = h(\theta), & \theta \in \Theta \\ \| \phi_0 \|_s = C_0. \end{cases}$$

and this completes the proof of the theorem.

It is readily seen that $\mathfrak{M}_s$ will consist of more than just ϕ_0 if and only if there exists a non-null functional on $\mathfrak{L}_r$ which vanishes on $\mathfrak{P}_0$. Our discussion in section 3 therefore enables us to assert the following.

COROLLARY 2-1. *$\mathfrak{M}_s$, when it is non-empty, consists of ϕ_0 alone if and only if $\mathfrak{P}_0$ is fundamental in $\mathfrak{L}_r$.*

A word is in order concerning the following two consequences of the boundedness of the measure ν: (i) if $\mathfrak{P}_0 \subset \mathfrak{L}_r$, then also $\mathfrak{P}_0 \subset \mathfrak{L}_{r'}$ for every $r' < r$; (ii) if $\phi \in \mathfrak{L}_s$, then also $\phi \in \mathfrak{L}_{s'}$ for every $s' < s$. Otherwise stated: (i') if $\mathfrak{P}_0$ satisfies the postulate of section 2 for the number r, it likewise satisfies this postulate for every (admissible) $r' < r$; (ii') if $\mathfrak{M}_s$ is non-empty, then $\mathfrak{M}_{s'}$ is non-empty for every $s' < s$. Regarding (i') we shall make only the obvious remark that although $\mathfrak{P}_0$ satisfies the postulate for every $r' < r$, there may be values of $r' < r$ such that no C for (5) exists; this will be exemplified in section 9. Where (ii') is concerned, it is clear that the non-emptiness of $\mathfrak{M}_s$ will not necessarily imply that $\mathfrak{P}_0 \subset \mathfrak{L}_{s'/s'-1}$ for every $s' < s$, even though for every such s' $\mathfrak{M}_{s'}$ is non-empty. If for every $\phi \in \Theta$ other than θ_0 we have $\pi_\theta \notin \mathfrak{L}_{s'/s'-1}$, for some particular $s' < s$, then we may have the situation in which there are elements in $\mathfrak{M}_{s'}$ with norms arbitrarily close to 0. However, this cannot be the case if (a) for some θ other than θ_0, $\pi_\theta \in \mathfrak{L}_{s'/s'-1}$, and (b) h does not vanish identically on Θ', the set of those θ for which $\pi_\theta \in \mathfrak{L}_{s'/s'-1}$. For, when these two conditions are satisfied, Theorem 2 applies to h as defined on Θ'; consequently there is a positive lower bound for the s'—norms of the unbiased estimates of h over Θ'. And since every element of $\mathfrak{M}_{s'}$ is, in particular, an unbiased estimate of h over Θ', it follows that the norms of those elements are bounded below by a positive number.

5. The case $s = \infty$ ($r = 1$). Let $\mathfrak{M}_\infty$ denote the class of essentially bounded (ν) unbiased estimates of h; and let bestness at θ_0 be defined with respect to the essential absolute suprema of the elements of this class. That is, the unbiased estimate ϕ_0, of h, is best at θ_0 if

$$\operatorname*{ess.\,sup.}_{x \in \Omega} |\phi_0(x)| < \infty,$$

and if, when ϕ is another unbiased estimate of h, we have

$$\operatorname*{ess.\,sup.}_{x \in \Omega} |\phi_0(x)| \leqq \operatorname*{ess.\,sup.}_{x \in \Omega} |\phi(x)|.$$

The fundamental postulate for the functions π_θ is, in this case, that $\mathfrak{P}_0 \subset \mathfrak{L}_1$.

Now, $\mathfrak{L}_\infty$, the space of essentially bounded, measurable (ν) functions on Ω, normed by ess. sup., is the space of linear functionals on $\mathfrak{L}_1$ [14, p. 338]. Examination of the proof of Theorem 2 will show that that proof goes through also in the present case in all but one detail: we cannot here in general prove the uniqueness of the best estimate. The proof of uniqueness breaks down since the equality

$$\text{ess. sup.} |\phi_0(x) + \phi_1(x)| = \text{ess. sup.} |\phi_0(x)| + \text{ess. sup.} |\phi_1(x)|$$

does not imply that ϕ_1 is a constant multiple of ϕ_0. Of course, if $\mathfrak{P}_0$ is fundamental in $\mathfrak{L}_1$, we have a fortiori the uniqueness of the best estimate.

The results for the case $s = \infty$ are then the following.

THEOREM 3. (i) *A necessary and sufficient condition that $\mathfrak{M}_\infty$ be non-empty is that there exist a constant C such that for every set of n functions $\pi_{\theta_1}, \pi_{\theta_2}, \cdots, \pi_{\theta_n}$, in $\mathfrak{P}_0$, and every set of n real numbers $a_1, a_2, \cdots, a_n$, we have, for every $n = 1, 2, \cdots$,*

$$\left| \sum_{i=1}^{n} a_i h(\theta_i) \right| \leqq C \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_1 .$$

(ii) *For every $\phi \in \mathfrak{M}_\infty$ we have $\| \phi \|_\infty \geqq C_0$, where C_0 is the g.l.b. of the set of admissible constants C above.*

(iii) *When $\mathfrak{M}_\infty$ is non-empty, it contains elements with norm equal to C_0. These are the best (at θ_0) unbiased estimates of h. When $\mathfrak{P}_0$ is not fundamental in $\mathfrak{L}_1$, there need not exist a unique best estimate.*

We close this section with the remark that Theorem 1 remains valid, as it stands, in the case $s = \infty$.

6. Particular lower bounds for the minimum s.a.c.m. In order to stress their significance in the statistical context, we shall give the statements of this section with the help of the symbol $\sigma_s(\phi)$ for the sth root of the s.a.c.m. of the unbiased estimate ϕ, of h. We have of course, the relation

$$\sigma_s(\phi) = \| \phi \|_s .$$

Now, one of the most important aspects of Theorem 2 is that it presents us immediately with an explicit evaluation of the minimum $\sigma_s(\phi)$ for all $\phi \in \mathfrak{M}_s$. We state the formula in the form of a theorem.

THEOREM 4. *Let $\mathfrak{R}$ denote the set of all real numbers. Then,*

$$\operatorname*{g\,l.b.}_{\phi \in \mathfrak{M}_s} \sigma_s(\phi) = \operatorname*{l.u.b.}_{\substack{\theta_1, \theta_2, \cdots, \theta_n \in \Theta \\ a_1, a_2, \cdots, a_n \in \mathfrak{R} \\ n = 1, 2, \cdots}} \frac{\left| \sum_{i=1}^{n} a_i h(\theta_i) \right|}{\left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r}$$

For brevity, let us set

$$\operatorname*{g.l.b.}_{\phi \in \mathfrak{M}_s} \sigma_s(\phi) = \sigma_s^{\min} .$$

Since this theorem expresses $\sigma_s^{\min}$ as the l.u.b. of an explicit set of numbers, it is clear that *the class of all lower bounds of $\sigma_s^{\min}$ is thereby thrown open to us.* It follows that, when $s = r = 2$ and our hypotheses on $\mathfrak{P}$ are fulfilled, the classical lower bounds of Cramér-Rao [3, p. 480] and Bhattacharyya [4, p. 3] are particularized consequences of Theorem 4. In the results that follow here we shall indicate the deduction of those classical bounds. We need not, however, restrict s.

For a moment, let us denote by $\pi(x)$ the function on Θ which assigns the value $\pi_\rho(x)$ to the point $\rho \in \Theta$, and let Θ be an interval on the real axis. Then we shall, below, write π'_θ for the function (when it exists) on Ω which assigns the

value $(d\pi(x)/d\rho)_{\rho=\theta}$ to $x \in \Omega$. Similarly, π''_θ for the function assigning the value $(d^2\pi(x)/d\rho^2)_{\rho=\theta}$ to x; and so on.

THEOREM 5. *Suppose the following conditions fulfilled:*

(i) $\Theta = \mathcal{I}$, *an interval on the real axis;*

(ii) h *is differentiable on* $\Theta' \subseteqq \mathcal{I}$,

(iii) *for each* $\theta \in \Theta'$, π'_θ *is defined almost everywhere* (ν), *and is an element of* $\mathfrak{L}_r$;

(iv) *for each* $\theta \in \Theta'$,

$$\lim_{\rho \to \theta} \left\| \frac{\pi_\rho - \pi_\theta}{\rho - \theta} - \pi'_\theta \right\|_r = 0.$$

Then, for any $m + n$ $(m, n = 1, 2, \cdots)$ *points* $\theta_1, \theta_2, \cdots, \theta_m$ *in* $\mathcal{I}$, *and* $\theta'_1, \theta'_2, \cdots, \theta'_n$ *in* Θ', *and any* $m + n$ *real numbers* $a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n$ *such that*

$$\left\| \sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \pi'_{\theta'_i} \right\|_r \neq 0,$$

we have

$$\sigma_\theta^{\min} \geqq \frac{\left| \sum_{i=1}^{m} a_i h(\theta_i) + \sum_{i=1}^{n} b_i h'(\theta'_i) \right|}{\left\| \sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \pi'_{\theta'_i} \right\|_r}. \tag{6}$$

The prime on the h in (6) denotes the derivative of h.

To prove this theorem, observe first that by virtue of Theorem 4, we may write

$$\sigma_\theta^{\min} \geqq \frac{\left| \sum_{i=1}^{m} a_i h(\theta_i) + \sum_{i=1}^{n} b_i \dfrac{h(\rho_i) - h(\theta'_i)}{\rho_i - \theta'_i} \right|}{\left\| \sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \dfrac{\pi_{\rho_i} - \pi_{\theta'_i}}{\rho_i - \theta'_i} \right\|_r}$$

for every set of points $\rho_1, \rho_2, \cdots, \rho_n$ in $\mathcal{I}$ such that the denominator of the right-hand side is defined and $\neq 0$. Therefore, also

$$\sigma_\theta^{\min} \geqq \lim_{\substack{\rho_i \to \theta_i \\ i=1,2,\cdots,n}} \frac{\left| \sum_{i=1}^{m} a_i h(\theta_i) + \sum_{i=1}^{n} b_i \dfrac{h(\rho_i) - h(\theta'_i)}{\rho_i - \theta'_i} \right|}{\left\| \sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \dfrac{\pi_{\rho_i} - \pi_{\theta'_i}}{\rho_i - \theta'_i} \right\|_r}. \tag{7}$$

Now, by condition (iv), the element

$$\sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \frac{\pi_{\rho_i} - \pi_{\theta'_i}}{\rho_i - \theta'_i}$$

of $\mathfrak{L}_r$ converges, in the strong sense in $\mathfrak{L}_r$, to

$$\sum_{i=1}^{m} a_i \pi_{\theta_i} + \sum_{i=1}^{n} b_i \pi'_{\theta'_i},$$

as $\rho_i \to \theta_i'$, $i = 1, 2, \cdots, n$. Consequently we have convergence of the norm; that is, the denominator of the right-hand side of (7) converges to the denominator of (6). (The latter is $\neq 0$, so that for all ρ_i sufficiently close to θ_i', $i = 1, 2, \cdots, n$, the ratios in (7) are defined.) There is no difficulty about the convergence of the numerator of (7) to that of (6). The theorem is thus proved.

COROLLARY 5-1. *Under the hypothesis of Theorem* 5, *we have, in particular, when* $\theta_0 \in \Theta'$ *and* $\| \overset{*}{\pi}{}'_{\theta_0} \|_r \neq 0$,

$$\sigma_i^{\min} \geqq \frac{|h'(\theta_0)|}{\| \overset{*}{\pi}{}'_{\theta_0} \|_r}. \tag{8}$$

If we denote by p the function on $\Omega \times \Theta$ which assigns the value $p_\theta(x)$ to the point (x, θ), and write (8) in the form

$$(\sigma_i^{\min})^r \geqq \frac{|h'(\theta_0)|^r}{\int_\Omega \left| \frac{\partial \log p}{\partial \theta} \right|_{\theta=\theta_0}^r p_{\theta_0}\, d\mu}, \tag{8'}$$

the generalization of the Cramér-Rao inequality afforded by (8) becomes evident.

Using the result and method of Theorem 5, we can establish the next in a hierarchy of theorems.

THEOREM 6. *Suppose the hypothesis of Theorem* 5 *satisfied, and the following condition fulfilled: for each* θ *in a non-empty subset* Θ'' *of* Θ', (i) $h''(\theta)$ (*the second derivative*) *exists and* (ii) π_θ'' *is defined almost everywhere* (ν), *is an element of* $\mathfrak{L}_r$, *and satisfies*

$$\lim_{\rho \to \theta} \left\| \frac{\pi_\rho' - \pi_\theta'}{\rho - \theta} - \pi_\theta'' \right\|_r = 0.$$

Then, for any $m + n + q (m, n, q = 1, 2, \cdots)$ *points* $\theta_1, \theta_2, \cdots, \theta_m$ *in* $\mathfrak{I}$, $\theta_1', \theta_2', \cdots, \theta_n'$ *in* Θ', *and* $\theta_1'', \theta_2'', \cdots, \theta_q''$ *in* Θ'', *and any* $m + n + q$ *real numbers* $a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n, c_1, c_2, \cdots, c_q$ *such that*

$$\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' + \sum_{i=1}^q c_i \pi_{\theta_i''}'' \right\|_r \neq 0,$$

we have

$$\sigma_i^{\min} \geqq \frac{\left| \sum_{i=1}^m a_i h(\theta_i) + \sum_{i=1}^n b_i h'(\theta_i') + \sum_{i=1}^q c_i h''(\theta_i'') \right|}{\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' + \sum_{i=1}^q c_i \pi_{\theta_i''}'' \right\|_r}.$$

Just as in the case of the previous theorem, we have here an immediate corollary.

COROLLARY 6-1. *Under the hypothesis of Theorem* 6, *we have in particular, when* $\theta_0 \in \Theta' \cdot \Theta''$,

$$\sigma_i^{\min} \geqq \frac{|bh'(\theta_0) + ch''(\theta_0)|}{\| b\pi_{\theta_0}' + c\pi_{\theta_0}'' \|_r}, \tag{9}$$

for any two real numbers, b and c, such that the denominator of the right-hand side does not vanish.

Consider (9) in the particular case $s = r = 2$. In this case, (9) may be written, explicitly,

$$(\sigma_2^{\min})^2 \geqq \frac{|bh'(\theta_0) + ch''(\theta_0)|^2}{\int_\Omega \frac{1}{p_{\theta_0}} \left(b \frac{\partial p}{\partial \theta} + c \frac{\partial^2 p}{\partial \theta^2} \right)_{\theta=\theta_0}^2 d\mu}. \tag{10}$$

In particular, (10) holds for values of b and c which maximize the right-hand side. And that maximum value is found, in the usual way, to be

$$J^{11}[h'(\theta_0)]^2 + 2J^{12}h'(\theta_0)h''(\theta_0) + J^{22}[h''(\theta_0)]^2,$$

where the matrix

$$\begin{pmatrix} J^{11} & J^{12} \\ J^{12} & J^{22} \end{pmatrix}$$

is the inverse of the matrix

$$\begin{pmatrix} \int_\Omega \frac{1}{p_\theta} \left(\frac{\partial p}{\partial \theta} \right)^2 d\mu & \int_\Omega \frac{1}{p_\theta} \frac{\partial p}{\partial \theta} \frac{\partial^2 p}{\partial \theta^2} d\mu \\ \int_\Omega \frac{1}{p_\theta} \frac{\partial p}{\partial \theta} \frac{\partial^2 p}{\partial \theta^2} d\mu & \int_\Omega \frac{1}{p_\theta} \left(\frac{\partial^2 p}{\partial \theta^2} \right)^2 d\mu \end{pmatrix}.$$

Thus, we have

$$(\sigma_2^{\min})^2 \geqq J^{11}[h'(\theta_0)]^2 + 2J^{12}h'(\theta_0)h''(\theta_0) + J^{22}[h''(\theta_0)]^2. \tag{11}$$

This is seen to be Bhattacharyya's result for the case of derivatives up to second order.

It is obvious how we extend Theorem 6 to obtain a similar result involving the functions π_θ, π_θ', π_θ'', $\cdots$, $\pi_\theta^{(n)}$, for any assigned n. And it is thereafter clear how, in the case $s = r = 2$, Bhattacharyya's general inequality may be deduced.

It is clear that we can proceed from Theorem 4, under suitable conditions, to lower bounds for $\sigma_s^{\min}$ which involve integrals of the functions $\pi(x)$ (and the corresponding integrals of h) as well as the derivatives of these functions.

In closing this section we note that all the above considerations apply equally to the case $s = \infty$.

7. Determination of the best estimate. We shall now prove the following theorem, which provides an explicit construction of the best (at θ_0) estimate of h. We repeat that s is now taken to be finite.

THEOREM 7. *Let $\mathfrak{M}_s$ be non-empty, and ϕ_0 be the best (at θ_0) unbiased estimate of h. Let $\{\theta_i^n, i = 1, 2, \cdots, k_n\}$, $n = 1, 2, \cdots$, be a sequence of (finite) sets of points of Θ, and $\{\alpha_i^n, i = 1, 2, \cdots, k_n\}$, $n = 1, 2, \cdots$, a sequence of sets of real numbers, such that*

$$\lim_{n\to\infty} \frac{\left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right|}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r} = C_0 = \| \phi_0 \|_s = \sigma_s^{\min}.$$

Then the functions ζ_n :

$$\zeta_n(x) = \frac{\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r^r} \cdot \left| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right|^{r/s} \operatorname{sgn} \left(\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right)$$

(*are elements of* $\mathfrak{L}_s$ *and*) *converge strongly in* $\mathfrak{L}_s$ *to* ϕ_0 .
The strong convergence here means precisely that

$$\lim_{n \to \infty} \int_\Omega | \zeta_n - \phi_0 |^s \, d\nu = 0.$$

Clearly, we may, with no loss in generality, assume the numbers α_i^n to be such that

$$\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r = 1, \; n = 1, 2, \cdots. \tag{12}$$

We shall suppose this to be the case throughout the proof. Then the essential property of the θ_i^n and the α_i^n is that

$$\lim_{n \to \infty} \left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right| = C_0. \tag{13}$$

And in this normalized situation, the functions ζ_n will be given by

$$\zeta_n(x) = \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \cdot \left| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right|^{r/s} \operatorname{sgn} \left(\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right). \tag{14}$$

That these functions are elements of $\mathfrak{L}_s$ is easily seen; in fact,

$$\| \zeta_n \|_s = \left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right|.$$

The proof of this theorem will consist mainly in the application of the following two lemmas.

LEMMA 2. *Let* $0 \neq \eta \in \mathfrak{L}_s$, *and* $\{\xi_n , n = 1, 2, \cdots\}$ *be a sequence of functions in* $\mathfrak{L}_r$ *such that*

$$\text{(i)} \quad \| \xi_n \|_r = 1, \qquad n = 1, 2, \cdots$$

$$\text{(ii)} \quad \lim_{n \to \infty} \int_\Omega \xi_n \eta \, d\nu = \| \eta \|_s .$$

Then ξ_n *converges strongly in* $\mathfrak{L}_r$ *to the function*

$$\xi_0 = \frac{1}{\| \eta \|_s^{s/r}} | \eta |^{s/r} \operatorname{sgn} \eta.$$

Let us observe first that

$$\int_\Omega \xi_0 \eta \, d\nu = \| \eta \|_s \tag{15}$$

and

$$|| \xi_0 || = 1.$$

Furthermore, ξ_0 is the unique element with norm $\leqq 1$ in $\mathfrak{L}_r$ having the property (15). For, if also,

$$\int_\Omega \xi_0' \eta \, d\nu = || \eta ||_s, \qquad || \xi_0' ||_r \leqq 1,$$

we then have

$$\int_\Omega \tfrac{1}{2}(\xi_0 + \xi_0') \cdot \eta \, d\nu = || \eta ||_s;$$

and from this,

$$\tfrac{1}{2} || \xi_0 + \xi_0' ||_r || \eta ||_s \geqq || \eta ||_s.$$

That is,

$$|| \xi_0 + \xi_0' ||_r \geqq 2 \geqq || \xi_0 ||_r + || \xi_0' ||_r.$$

From this, and (Minkowski)

$$|| \xi_0 + \xi_0' ||_r \leqq || \xi_0 ||_r + || \xi_0' ||_r,$$

we have

$$|| \xi_0 + \xi_0' ||_r = || \xi_0 ||_r + || \xi_0' ||_r.$$

Therefore, for some $a > 0$, $\xi_0' = a\xi_0$. But we must have $a = 1$ if ξ_0 and ξ_0' are both to satisfy (15), as assumed. Hence $\xi_0' = \xi_0$.

Now consider the sequence $\{\xi_n\}$. Choose a sub-sequence $\{\xi_{n_i}\}$ that converges weakly to, say, ξ'. Then $|| \xi' ||_r \leqq 1$. We have

$$\int_\Omega \xi' \eta \, d\nu = \lim_{i \to \infty} \int_\Omega \xi_{n_i} \eta \, d\nu = || \eta ||_s.$$

Hence, $\xi' = \xi_0$. And since $1 = || \xi_{n_i} ||_r \to 1 = || \xi_0 ||_r$, it follows that ξ_{n_i} converges strongly to ξ_0 (cf. [13, p. 139, section 3]).

Suppose there is a subsequence $\{\xi_{n_i}'\}$ of $\{\xi_n\}$ such that

$$|| \xi_{n_i}' - \xi_0 || > \delta > 0, \qquad i = 1, 2, \cdots.$$

We have, nonetheless, for this subsequence, the hypotheses of our lemma satisfied. We can therefore apply the argument of the previous paragraph to extract a subsequence of $\{\xi_{n_i}'\}$, which converges strongly to ξ_0 This is in obvious contradiction to the above δ-assumption, and the lemma is hereby proved.

LEMMA 3. *Lemma 2 remains true with the roles of $\mathfrak{L}_r$ and $\mathfrak{L}_s$ interchanged.*
This is obvious

Returning now to the proof of Theorem 7, let us first, for the sake of brevity,

introduce the notation:

$$c_n = \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n),$$

$$\gamma_n = \operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)\right),$$

$$\psi_n = \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}.$$

From

$$\int_\Omega \phi_0 \pi_\theta \, d\nu = h(\theta), \qquad \theta \in \Theta,$$

we easily obtain

$$\int_\Omega \phi_0 \psi_n \, d\nu = c_n, \qquad n = 1, 2, \cdots,$$

which we may write

$$\int_\Omega \phi_0 \cdot \gamma_n \psi_n \, d\nu = |c_n|, \qquad n = 1, 2, \cdots.$$

Since $|c_n| \to ||\phi_0||_s$ (cf. (13)) and $||\gamma_n \psi_n||_r = 1$, $n = 1, 2, \cdots$, (cf. (12)), we have, by Lemma 2, that $\gamma_n \psi_n$ converges strongly to

$$\psi_0 = \frac{1}{C_0^{s/r}} |\phi_0|^{s/r} \operatorname{sgn} \phi_0. \tag{16}$$

The functions (cf. (14))

$$\zeta_n = c_n |\psi_n|^{r/s} \operatorname{sgn} \psi_n$$

obviously satisfy

$$\int_\Omega \zeta_n \cdot \gamma_n \psi_n \, d\nu = |c_n|, \qquad n = 1, 2, \cdots$$

And from this we conclude that

$$\lim_{n \to \infty} \int_\Omega \zeta_n \psi_0 \, d\nu = C_0,$$

or

$$\lim_{n \to \infty} \int_\Omega \frac{\zeta_n}{|c_n|} \psi_0 \, d\nu = 1 = ||\psi_0||_r.$$

We may apply Lemma 3 to this result, since $||\zeta_n / |c_n|\,||_r = 1$, $n = 1, 2, \cdots$. And we thereby conclude that $\zeta_n / |c_n|$ converges strongly to

$$|\psi_0|^{r/s} \operatorname{sgn} \psi_0,$$

which, on substituting from the definition (16) of ψ_0, we find to be just

$$\frac{\phi_0}{C_0}.$$

Since $|c_n| \to C_0$, it follows immediately that ζ_n converges strongly to ϕ_0, and the theorem is proved.

The following corollary is actually of greater use in applications than Theorem 7 itself, for the reason that it leaves no doubt about the form of $\lim \zeta_n$ (i.e, ϕ_0) when we know explicitly the form of $\lim \gamma_n \psi_n$.

COROLLARY 7-1. *Assume the hypothesis of Theorem 7. Then the functions*

$$\frac{\operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)\right)}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r} \cdot \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}$$

converge strongly, in $\mathfrak{L}_r$, *to a function* ψ_0, *and*

$$\phi_0 = C_0 |\psi_0|^{r/s} \operatorname{sgn} \psi_0.$$

This is clear from the proof of the theorem.

By way of illustrating the application of these results, we shall prove the following theorem

THEOREM 8. *Assume the hypothesis of Theorem 5. And, further, let the equality sign hold in* (8). *Then,*

$$\phi_0(x) = \frac{h'(\theta_0)}{\| \pi'_{\theta_0} \|_r^r} \cdot |\pi'_{\theta_0}(x)|^{r/s} \operatorname{sgn} \pi'_{\theta_0}(x).$$

Since (8) is an equality, we may under the hypothesis of Theorem 5, consider that we have

$$C_0 = \lim_{n \to \infty} \frac{\left| \frac{1}{\rho_n - \theta_0} h(\rho_n) - \frac{1}{\rho_n - \theta_0} h(\theta_0) \right|}{\left\| \frac{1}{\rho_n - \theta_0} \pi_{\rho_n} - \frac{1}{\rho_n - \theta_0} \pi_{\theta_0} \right\|_r}. \tag{17}$$

where $\{\rho_n\}$ is a sequence in $\mathcal{I}$ converging to θ_0. The numerator of the right-hand side of (17), sans the vertical bars, converges to $h'(\theta_0)$ (which is $\neq 0$, since $C_0 \neq 0$); hence, for all sufficiently large n, that expression has the signum of $h'(\theta_0)$. The functions whose norms appear in the denominator of (17) we know to converge strongly in $\mathfrak{L}_r$ to π'_{θ_0} (by the hypothesis of Theorem 5). Hence, for this case, the function ψ_0 of Corollary 7-1 is

$$\psi_0 = \frac{\operatorname{sgn} h'(\theta_0)}{\| \pi'_{\theta_0} \|_r} \pi'_{\theta_0}.$$

Therefore, by the same corollary,

$$\phi_0(x) = \frac{|h'(\theta_0)|}{\| \pi'_{\theta_0} \|_r} \cdot \left| \frac{\operatorname{sgn} h'(\theta_0)}{\| \pi'_{\theta_0} \|_r} \pi'_{\theta_0} \right|^{r/s}$$

$$\cdot \operatorname{sgn} h'(\theta_0) \cdot \operatorname{sgn} \pi'_{\theta_0}(x)$$

$$= \frac{h'(\theta_0)}{\| \pi'_{\theta_0} \|_r^r} \mid \pi'_{\theta_0}(x) \mid^{r/s} \operatorname{sgn} \pi'_{\theta_0}(x).$$

And this is the result asserted in the theorem.

The reader will have no difficulty in establishing, in the exact pattern of the preceding proof, the following.

THEOREM 9. *Assume the hypothesis of Theorem* 6. *And, further, let the equality sign hold in* (9) *for* $b = b_0$, $c = c_0$ · [6] *Then,*

$$\phi_0(x) = \frac{b_0 h'(\theta_0) + c_0 h''(\theta_0)}{\| b_0 \pi'_{\theta_0} + c_0 \pi''_{\theta_0} \|_r^r} \cdot \mid b_0 \pi'_{\theta_0}(x) + c_0 \pi''_{\theta_0}(x) \mid^{r/s} \cdot \operatorname{sgn} (b_0 \pi'_{\theta_0}(x) + c_0 \pi''_{\theta_0}(x)).$$

It is evident that results of the type in these theorems may be built up as well with integrals over the parameter space.

A question of considerable practical importance is that of the rapidity of convergence of the ζ_n to ϕ_0. An answer to this question, on the level of generality we are maintaining in this study, consists in relating this convergence to that of the $\mid c_n \mid$ to C_0. In the case $s = r = 2$, the answer is immediate and exact:

$$\| \zeta_n - \phi_0 \|_2^2 = \int_\Omega (\zeta_n - \phi_0)^2 \, d\nu$$

$$= \int_\Omega \zeta_n^2 \, d\nu - 2 \int_\Omega \phi_0 \zeta_n \, d\nu + \int_\Omega \phi_0^2 \, d\nu$$

$$= \mid c_n \mid^2 - 2 \mid c_n \mid^2 + C_0^2$$

$$= C_0^2 - \mid c_n \mid^2.$$

Thus, if one unbiased estimate is known, it provides, since its norm is $\geqq C_0$, an upper bound for $\| \zeta_n - \phi_0 \|_2$ The same is true in the general case (any s) once we have established an upper bound, depending on C_0 and $\mid c_n \mid$, for $\| \zeta_n - \phi_0 \|_s$. But in the general case, a good upper bound does not seem to be so close at hand. There are indications of the direction in which one must proceed, and we hope to draw some significant results out of these before long.

8. The case $s = r = 2$. The particular aspects of this case (where bestness of an estimate has reference to its *variance*), which arise out of the coincidence of $\mathfrak{L}_r$ and $\mathfrak{L}_s$, merit some discussion. We shall denote the inner product, $\int_\Omega \xi \eta \, d\nu$, of two functions ξ and η in $\mathfrak{L}_2$, as usual by (ξ, η). Let $\{\mathfrak{P}_0\}$ denote the closed linear manifold in $\mathfrak{L}_2$ spanned by the π_θ.

THEOREM 10. *Let $\mathfrak{M}_2$ be non-empty. Then ϕ_0 is the unique element of $\mathfrak{M}_2$ which lies in $\{\mathfrak{P}_0\}$.*

[6] In the case $s = 2$, b_0 and c_0 are the values which render (11) an equality.

To begin with it is clear that the functions ζ_n of Theorem 7, in the present case $s = r = 2$, are all elements of $[\mathfrak{P}_0]$, the linear manifold spanned by the π_θ. Hence, since ϕ_0 is the strong limit of these elements, $\phi_0 \in \{\mathfrak{P}_0\}$.

Now suppose also $\phi_1 \in \mathfrak{M}_2$, $\phi_1 \in \{\mathfrak{P}_0\}$. Then, from

$$(\phi_0 , \pi_\theta) = h(\theta), \qquad \theta \in \Theta,$$

$$(\phi_1 , \pi_\theta) = h(\theta), \qquad \theta \in \Theta,$$

we have

$$(\phi_1 - \phi_0 , \pi_\theta) = 0, \qquad \theta \in \Theta,$$

and, by continuity of the inner product,

$$(\phi_1 - \phi_0 , \xi) = 0, \qquad \xi \in \{\mathfrak{P}_0\};$$

that is, $\phi_1 - \phi_0 \in \{\mathfrak{P}_0\}^\perp$. But, from $\phi_0 \in \{\mathfrak{P}_0\}$ and $\phi_1 \in \{\mathfrak{P}_0\}$ it follows that $\phi_1 - \phi_0 \in \{\mathfrak{P}_0\}$. Hence $\phi_1 - \phi_0 = 0$, and this proves the exclusiveness of the property for ϕ_0.

Another characterization of ϕ_0 is given by the following corollary

COROLLARY 10-1. *If $\mathfrak{M}_2$ is non-empty, then ϕ_0 is the unique element of $\mathfrak{M}_2$ which satisfies the system of equations in ξ : $(\phi, \xi) = \| \xi \|_2^2$, $\phi \in \mathfrak{M}_2$.*

To see that ϕ_0 has the asserted property, let ϕ be any element of $\mathfrak{M}_2$, and set $\phi = \xi + \eta$, with $\xi \in \{\mathfrak{P}_0\}$ and $\eta \in \{\mathfrak{P}_0\}^\perp$. From

$$(\xi, \pi_\theta) = (\xi + \eta, \pi_\theta) = (\phi, \pi_\theta) = h(\theta),$$

it follows that $\xi \in \mathfrak{M}_2$. Hence $\xi = \phi_0$. And so,

$$(\phi, \phi_0) = (\phi_0 + \eta, \phi_0) = \| \phi_0 \|_2^2$$

If $\phi_1 \in \mathfrak{M}_2$ has this property also, then both

$$(\phi_1 , \phi_0) = \| \phi_0 \|_2^2$$

and

$$(\phi_0 , \phi_1) = \| \phi_1 \|_2^2 ;$$

and therefore

$$\| \phi_1 \|_2 = \| \phi_0 \|_2 .$$

This proves $\phi_1 = \phi_0$, and so the corollary

9. An example. Let Ω be Euclidean n-space, $x = (x_1 , x_2 , \cdots , x_n)$; μ, Lebesgue measure; Θ, the set of real numbers; and

$$p_\theta(x) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (x_i - \theta)^2\right\}.$$

And finally, let $\theta_0 = 0$. Then

$$\pi_\theta(x) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (-2\theta x_i + \theta^2)\right\}.$$

If $0 < b < \frac{1}{2}$, and we define

$$\phi_1(x) = (1 - 2b)^{n/2} \exp\left\{b \sum_{i=1}^{n} x_i^2\right\} - 1,$$

we have, for each θ,

$$\int_a \phi_1(x) p_\theta(x)\, d\mu = \exp\left\{\frac{nb}{1 - 2b}\theta^2\right\} - 1.$$

Thus, ϕ_1 is an unbiased estimate of the function h:

$$h(\theta) = \exp\left\{\frac{nb}{1 - 2b}\theta^2\right\} - 1.$$

If we examine

$$\| \phi_1 \|_s^s = \frac{1}{(2\pi)^{ns/2}} \int_a \left| (1 - 2b)^{n/2} \exp\left\{b \sum_{i=1}^{n} x_i^2\right\} - 1 \right|^s \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2\right\} d\mu;$$

we find that this integral converges only for $s < 1/2b$. Shifting the emphasis, we may state: *for the function h, defined by*

$$h(\theta) = e^{\alpha\theta^2} - 1, \qquad \alpha > 0,$$

there exists an unbiased estimate with finite sth moment at $\theta = 0$, for each

$$s < \frac{n + 2\alpha}{2\alpha}.$$

Next, observe that

$$\| \pi_\theta \|_r^r = \frac{1}{(2\pi)^{nr/2}} \int_a \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i^2 - 2r\theta x_i + r\theta^2)\right\} d\mu$$

$$= \exp\{\tfrac{1}{2} nr(r - 1)\theta^2\},$$

so that the π_θ are elements of $\mathfrak{L}_r$ for each $r > 1$. The ratio

$$\frac{|h(\theta)|}{\| \pi_\theta \|_r} = (e^{\alpha\theta^2} - 1) \exp\{-\tfrac{1}{2} n(r - 1)\theta^2\}$$

is seen to diverge as $\theta \to \infty$, if

$$\tfrac{1}{2} n(r - 1) < \alpha.$$

Hence, by Theorem 2, there exists no unbiased estimate of h belonging to $\mathfrak{L}_s$ for a value of s such that the number

$$r = \frac{s}{s - 1}$$

satisfies the inequality just above; that is, for a value of s greater than

$$\frac{n + 2\alpha}{2\alpha}.$$

Otherwise stated: *there exists no unbiased estimate of h with finite* sth *moment at* $\theta = 0$, *for*

$$s > \frac{n + 2\alpha}{2\alpha}.$$

It is most likely true that this last statement holds, in general, with

$$s \geqq \frac{n + 2\alpha}{2\alpha}.$$

We shall consider here only the case

$$\frac{n + 2\alpha}{2\alpha} = 2;$$

and since the analysis is the same for every pair n, α satisfying this equality, we treat the particular case of

$$n = 1, \qquad \alpha = \tfrac{1}{2}.$$

Thus, we shall show: *for* $n = 1$, *there exists no unbiased estimate of* h_2,

$$h_2(\theta) = e^{\frac{1}{2}\theta^2} - 1,$$

with finite variance at $\theta = 0$.

We must show that the ratios

$$\frac{\left|\sum_{i=1}^{m} a_i(e^{\frac{1}{2}\theta_i^2} - 1)\right|}{\left\|\sum_{i=1}^{m} a_i \pi_{\theta_i}\right\|_2}$$

are not bounded for all choices of m (distinct) θ_i's, and all sets of m real numbers a_i, and all m. This is clearly equivalent to showing the same for the ratios

$$Q(m, a_i, \theta_i) = \frac{\left|\sum_{i=1}^{m} a_i(1 - e^{-\frac{1}{2}\theta_i^2})\right|}{\left\|\sum_{i=1}^{m} a_i e^{-\frac{1}{2}\theta_i^2} \pi_{\theta_i}\right\|_2}.$$

Now we find, by direct computation,

$$\left\|\sum_{i=1}^{m} a_i e^{-\frac{1}{2}\theta_i^2} \pi_{\theta_i}\right\|_2^2 = \sum_{i,j=1}^{m} e^{-\frac{1}{2}(\theta_i - \theta_j)^2} a_i a_j.$$

And the solution of the familiar extremum problem:

$$\sup_{(a_i)} \left|\sum_{i=1}^{m} a_i(1 - e^{-\frac{1}{2}\theta_i^2})\right| \quad \text{subject to} \quad \sum_{i,j=1}^{m} e^{-\frac{1}{2}(\theta_i - \theta_j)^2} a_i a_j = 1$$

yields

$$\sup_{(a_i)} Q^2(m, a_i, \theta_i) = \sum_{i,j=1}^{m} v_{ij}(1 - e^{-\frac{1}{2}\theta_i^2})(1 - e^{-\frac{1}{2}\theta_j^2}),$$

where the matrix

$$V = (v_{ij}), \qquad i, j = 1, 2, \cdots, m,$$

is the inverse of the matrix

$$U = (e^{-\frac{1}{2}(\theta_i - \theta_j)^2}), \qquad i, j = 1, 2, \cdots, m.$$

We now take, in particular,

$$\theta_i = it, \qquad i = 1, 2, \cdots, m,$$

where t is a positive number. Clearly, there exists a number t_0 such that for $t > t_0$,

$$U(t) = (e^{-\frac{1}{2}(i-j)^2 t^2})$$

is non-singular. Also,

$$\lim_{t \to \infty} U(t) = I,$$

the identity matrix. Then, for $t > t_0$, $V = U^{-1}$ is a continuous function of U, so that

$$\lim_{t \to \infty} V(t) = (\lim_{t \to \infty} U(t))^{-1} = I.$$

Hence,

$$\lim_{t \to \infty} v_{ij}(t) = \delta_{ij}.$$

It follows that

$$\lim_{t \to \infty} \sup_{(a_i)} Q^2(m, a_i, it) = m,$$

and therefore,

$$\sup_{(a_i, \theta_i)} Q^2(m, a_i, \theta_i) \geqq m.$$

(A simple argument on the characteristic values of U shows that there is actually equality here.) This result gives the unboundedness of the ratios Q_j and our proposition is proved, by virtue of Theorem 2.

APPENDIX

The spaces $\mathfrak{L}_r$ and $\mathfrak{L}_s$ are instances of a Banach space over the reals; that is, a complete, normed, linear vector space, closed under multiplication by real numbers. That the space, say $\mathfrak{B}$, is normed is to say that there is a non-negative, real-valued function, $\| \ \|$, defined on $\mathfrak{B}$, with the properties:

$$\| \xi \| = 0 \quad \text{if and only if } \xi \text{ is the null vector,}$$
$$\| a\xi \| = | a | \cdot \| \xi \|,$$
$$\| \xi + \eta \| \leqq \| \xi \| + \| \eta \|;$$

where $\xi, \eta \in \mathfrak{B}$ and a is real. The number $\| \xi \|$ is called the *norm* of ξ.

The function $\| \xi - \eta \|$ on pairs ξ, η of vectors is a distance function in the usual sense. With it, *strong convergence* (or simply *convergence*) is defined in $\mathfrak{B}$: ξ_n *converges strongly* to ξ when $\lim_{n\to\infty} \| \xi_n - \xi \| = 0$. In symbols: $\xi_n \to \xi$ or $\lim \xi_n = \xi$. The usual set-theoretic notions are now defined in the obvious way; e.g., limit point of a set, closed set, etc That the space $\mathfrak{B}$ is complete means that every sequence $\{\xi_n\}$ satisfying $\lim_{m,n\to\infty} \| \xi_m - \xi_n \| = 0$ converges to a (unique) element $\xi \in \mathfrak{B}$.

A *linear manifold* $\mathfrak{M}$ in $\mathfrak{B}$ is a subset of $\mathfrak{B}$ with the property that for any two elements ξ, $\eta \in \mathfrak{M}$ and any two real numbers a, b, we have also $a\xi + b\eta \in \mathfrak{M}$. A *closed linear manifold* is a linear manifold that is closed in the set-theoretic sense. If S is any subset of $\mathfrak{B}$, then the set, $[S]$, of all finite linear combinations of elements of S is a linear manifold, it is the *linear manifold spanned by* S. The closure of $[S]$, denoted by $\{S\}$, is called the *closed linear manifold spanned by* S. In general, $[S]$ is a proper subset of $\{S\}$. A set $S \subseteq \mathfrak{B}$ is called *fundamental* when $\{S\} = \mathfrak{B}$.

A *linear functional*, G, on $\mathfrak{B}$ is a real-valued function with the property that for any two elements ξ, $\eta \in \mathfrak{B}$ and any two real numbers a, b, we have $G(a\xi + b\eta) = aG(\xi) + bG(\eta)$. The linear functional G is said to be *bounded* when the number

$$\| G \| = \text{l.u.b.}_{\| \xi \| \neq 0} \frac{| G(\xi) |}{\| \xi \|}$$

is finite. $\| G \|$ is called the *norm* of G. (Throughout the text of the paper, the qualification "bounded" has been understood in all references to linear functionals). If we define the sum of two linear functionals F and G by $(F + G)(\xi) = F(\xi) + G(\xi)$, and make the other requisite definitions in the obvious way, we find that the bounded linear functionals on $\mathfrak{B}$ form a linear vector space over the reals. The function $\| \|$ on the bounded linear functionals, which we have already called a norm, is in fact a norm in the Banach space sense. This vector space, so normed, is readily shown to be complete Hence it is a Banach space—usually called the conjugate space to $\mathfrak{B}$. It is this space we have referred to in the text as the *space of linear functionals on* $\mathfrak{B}$.

If a sequence $\{\xi_n\}$ of elements of $\mathfrak{B}$ has the property that $\lim_{n\to\infty} G(\xi_n) = G(\xi)$ for every bounded linear functional G, then ξ_n is said to *converge weakly to* ξ. If, of the sequence $\{\xi_n\}$, we know only that $\lim_{n\to\infty} G(\xi_n)$ exists for every bounded linear functional, we say simply that the sequence is weakly convergent. The space $\mathfrak{B}$ is called weakly complete if every weakly convergent sequence converges weakly to a limit. The spaces $\mathfrak{L}_r$, $r \geqq 1$ are weakly complete $\mathfrak{B}$ is said to be *weakly compact* if every bounded set $S \subset \mathfrak{B}$ contains a weakly convergent sequence. That S is "bounded" means $\text{l.u.b.}_{\xi \in S} \| \xi \| < \infty$.

A real Hilbert space $\mathfrak{H}$ is a real Banach space on which there is defined an

inner product; that is, a function (ξ, η) on pairs of elements ξ, η, with the properties

$$(\xi, \eta) = (\eta, \xi),$$

$$(a\xi, \eta) = a(\xi, \eta),$$

$$(\xi + \zeta, \eta) = (\xi, \eta) + (\zeta, \eta),$$

$$\| \xi \|^2 = (\xi, \xi).$$

The inner product is a *continuous* function of both its arguments; i.e., $\lim \xi_m = \xi$ and $\lim \eta_n = \eta$ imply $\lim (\xi_m, \eta_n) = (\xi, \eta)$. The space $\mathfrak{L}_2$ in the text is a Hilbert space when we take $(\xi, \eta) = \int_\Omega \xi\eta \, d\nu$. Two elements ξ, η which are such that $(\xi, \eta) = 0$ are said to be orthogonal. If S is any set in $\mathfrak{H}$, then the set of elements of $\mathfrak{H}$ each of which is orthogonal to every element of S is called the *orthocomplement* of S, and is denoted by $S^\perp$.

For further elaboration the reader is referred to [13] and [19].

REFERENCES

[1] D. Blackwell and M. A. Girshick, "A lower bound for the variance of some unbiased sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 277–280.

[2] R. A. Fisher, "Theory of statistical estimation," *Camb. Phil. Soc. Proc.*, Vol. 22, (1925), pp. 700–725.

[3] H. Cramér, *Mathematical Methods of Statistics*, Princeton Press, Princeton, 1946.

[4] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1–14.

[5] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[6] J. Neyman, "Su un teorema concernente le cosidette statistiche sufficienti," *Giornale dell'Istituto Italiano degli Attuari*, Vol. 6 (1935), pp. 320–334.

[7] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[8] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Aktuar. tids.*, Vol. 29 (1946), pp. 85–94.

[9] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation (cont'd)," *Sankhyā*, Vol. 8 (1947), pp. 201–218.

[10] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 215–230.

[11] F. Riesz, *Les Systèmes d'Équations Linéaires à une Infinité d'Inconnues*, Gauthier-Villars, Paris, 1913.

[12] F. Riesz, "Untersuchungen über Systeme integrierbare Funktionen," *Math. Annalen*, Vol. 69 (1910), pp. 449–497.

[13] S. Banach, *Théorie des Opérations Linéaires*, Garasinski, Warsaw, 1932.

[14] N. Dunford, "Uniformity in linear spaces," *Am. Math. Soc. Trans.*, Vol. 44 (1938), pp. 305–356.

[15] M. H. Steinhaus, "Additive und stetige funktionaloperationen," *Math. Zeits.*, Vol. 5 (1918), pp. 186–221.

[16] F. Riesz, "Sur une espèce de géométrie analytique des systèmes de fonctions sommables," *Comptes Rendus*, Vol. 144 (1907), pp. 1409–1411.
[17] M. Fréchet, "Sur les ensembles de fonctions et les opérations linéaires," *Comptes Rendus*, Vol 144 (1907), pp. 1414–1416.
[18] S. Saks, *Theory of the Integral*, Stechert, New York, 1937
[19] J. von Neumann, *Functional Operators*, (Mimeographed notes) Princeton, 1935
[20] G. R. Seth, "On the variance of estimates," *Annals of Math. Stat*, Vol. 20 (1949), pp. 1–27.
[21] B. J Pettis, "A note on regular Banach spaces," *Am. Math. Soc Bull.*, Vol 44 (1938), pp. 420–428.

A SEQUENTIAL DECISION PROCEDURE FOR CHOOSING ONE OF THREE HYPOTHESES CONCERNING THE UNKNOWN MEAN OF A NORMAL DISTRIBUTION

By Milton Sobel and Abraham Wald[1]

Columbia University

1. Introduction. In this paper a multi-decision problem is investigated from a sequential viewpoint and compared with the best non-sequential procedure available. Multi-decision problems occur often in practice but methods to deal with such problems are not yet sufficiently developed.

The problem under consideration here is a 3-decision problem: Given a chance variable which is normally distributed with known variance σ^2, but unknown mean θ, and given two real numbers $a_1 < a_2$, the problem is to choose one of the three mutually exclusive and exhaustive hypotheses

$$H_1 : \theta < a_1 \qquad H_2 : a_1 \leqq \theta \leqq a_2 \qquad H_3 : \theta > a_2 .$$

In order to select a proper sequential decision procedure, the parameter space is subdivided into 5 mutually exclusive and exhaustive zones in the following manner. Around a_1 there exists an interval (θ_1 , θ_2) in which we have no strong preference between H_1 and H_2 but prefer (strongly) to reject H_3. Around a_2 there exists an interval (θ_3 , θ_4) in which we have no strong preference between H_2 or H_3 but prefer (strongly) to reject H_1. For $\theta \leqq \theta_1$ we prefer to accept H_1. For $\theta_2 \leqq \theta \leqq \theta_3$ we prefer to accept H_2. For $\theta \geqq \theta_4$ we prefer to accept H_3.

The intervals (θ_1 , θ_2) and (θ_3 , θ_4) will be called indifference zones. The determination of these indifference zones is not a statistical problem but should be made on practical considerations concerning the consequences of a wrong decision.

In accordance with the above we define a wrong decision in the following way. For $\theta \leqq \theta_1$, acceptance of H_2 or H_3 is wrong. For $\theta_1 < \theta < \theta_2$ acceptance of H_3 is wrong. For $\theta_2 \leqq \theta \leqq \theta_3$, acceptance of H_1 or H_3 is wrong. For $\theta_3 < \theta < \theta_4$, acceptance of H_1 is wrong. For $\theta \geqq \theta_4$, acceptance of H_1 or H_2 is wrong.

The requirements on our decision procedure necessary to limit the probability of a wrong decision are investigated. Two cases are considered.

Case 1: Prob. of a wrong decision $\leqq \gamma$ for all θ.

Case 2:
- Prob. of a wrong decision $\leqq \gamma_1$ for $\theta \leqq \theta_1$,
- Prob. of a wrong decision $\leqq \gamma_2$ for $\theta_1 < \theta < \theta_4$,
- Prob. of a wrong decision $\leqq \gamma_3$ for $\theta \geqq \theta_4$.

The decision procedure discussed in the present paper is not an optimum procedure since, as will be seen later, the final decision at the termination of

[1] Work done under the sponsorship of the Office of Naval Research.

experimentation is not in every case a function of only "the sample mean of *all* the observations", although the sample mean is a sufficient statistic for θ. Although the procedure considered is not optimal it is suggested for the following reasons:

1. The decision procedure can be carried out simply In fact tables can be constructed before experimentation starts that render the procedure completely mechanical.

2. The derivation of the operating characteristic (OC) function, neglecting the excess of the cumulative sum over the boundary, is accomplished with little difficulty. In general, for other multi-decision problems it is unknown how to obtain the OC function.

3. It is believed that the loss of efficiency is not serious, i.e, the suggested sequential procedure is not far from being optimum. In this connection a non-sequential procedure is compared with this sequential procedure. The results show that, for the same maximum probability of making a wrong decision, the sequential procedure requires on the average substantially fewer observations to reach a final decision. In fact, for Case 1 noted above, if $.008 < \gamma < .1$, and if certain symmetrical features are assumed, then the fixed number of observations required by the non-sequential method is greater than the maximum of the average sample number (ASN) function taken over all values of θ.

It was found necessary in the course of the investigation to put an upper bound on the quantity $\frac{\theta_4 - \theta_3}{a_2 - a_1}$ in order that the methods used to obtain upper and lower bounds for the ASN function should give close results. This restriction, however, is likely to be satisfied in practical applications

All formulas for ASN and OC functions which will be used in this paper will be approximation formulas neglecting the excess of the cumulative sum over the boundaries Nevertheless, equality signs will be used in these formulas, except when additional approximations are involved.

2. Description of the Decision Procedure.[2] We shall assume that the indifference zones described above have the following properties

$$\text{(i)}\quad \theta_1 < a_1 < \theta_2 \leqq \theta_3 < a_2 < \theta_4$$

$$\text{(ii)}\quad \theta_1 + \theta_2 = 2a_1\ ; \quad \theta_3 + \theta_4 = 2a_2$$

$$\text{(iii)}\quad \theta_2 - \theta_1 = \theta_4 - \theta_3 = \Delta \text{ (say)}.$$

[2] A similar decision procedure was used by P. Armitage [2] as an alternative to the sequential t test (with 2-sided alternatives) The form used there is more restricted as he considers only the case $\theta_2 = \theta_3$. Essential inequalities on the OC function are pointed out but no attempt is made to determine the complete OC and ASN functions. A closely related but somewhat different procedure for dealing with a trichotomy was suggested by Milton Friedman while he was a member of the Statistical Research Group of Columbia University As far as the authors are aware, no results were obtained concerning the OC and ASN functions of Friedman's procedure.

Let R_1 denote the Sequential Probability Ratio Test for testing the hypothesis that $\theta = \theta_1$ against the hypothesis that $\theta = \theta_2$. We assume for the present that either the proper constants A, B in the probability ratio test are given or that they are approximated from given α, β by the relations

$$A \sim \frac{1-\beta}{\alpha} \qquad B \sim \frac{\beta}{1-\alpha}.$$

Here α and β are upper bounds on the probabilities of first and second types of errors, respectively.

Let R_2 represent the S.P.R.T. for testing the hypothesis that $\theta = \theta_3$ against the alternative that $\theta = \theta_4$. For this test we assume that (α, β, A, B) are replaced by $(\hat{\alpha}, \hat{\beta}, \hat{A}, \hat{B})$ and as above that either $\hat{A}$ and $\hat{B}$ are given or that they are approximated from given $\hat{\alpha}$, $\hat{\beta}$.

The decision procedure is carried out as follows:

Both R_1 and R_2 are computed at each stage of the inspection until

Either: One ratio leads to a decision to stop before the other. Then the former is no longer computed and the latter is continued until it leads to a decision to stop.

Or: Both R_1 and R_2 lead to a decision to stop at the same stage. In this event both computations are discontinued.

The following table gives the rule R for the decisions to be made corresponding to all possible outcomes of R_1 and R_2.

	R_1		R_2		R
If	accepts θ_1	and	accepts θ_3	then	accepts H_1
If	accepts θ_2	and	accepts θ_3	then	accepts H_2
If	accepts θ_2	and	accepts θ_4	then	accepts H_3

We shall show that acceptance of both θ_1 and θ_4 is impossible when $(\hat{A}, \hat{B}) = (A, B)$. For this purpose we need the acceptance number and rejection number formulas. (See page 119 of [1]).

	Acceptance Number		*Rejection Number*
R_1:	$\frac{\sigma^2}{\Delta} \log B + a_1 n$	$< \sum_{\alpha=1}^{n} x_\alpha <$	$\frac{\sigma^2}{\Delta} \log A + a_1 n$
R_2:	$\frac{\sigma^2}{\Delta} \log B + a_2 n$	$< \sum_{\alpha=1}^{n} x_\alpha <$	$\frac{\sigma^2}{\Delta} \log A + a_2 n.$

We shall assume *for convenience* that "between observations" R_1 is tested before R_2 and let the term "initial decision" refer to the first decision made.

Assume θ_1 and θ_4 are both accepted. Then if θ_1 is accepted initially at the mth stage

$$\sum_{\alpha=1}^{m} x_\alpha \leqq \frac{\sigma^2}{\Delta} \log B + a_1 m.$$

Since

$$\frac{\sigma^2}{\Delta} \log B + a_1 m < \frac{\sigma^2}{\Delta} \log B + a_2 m$$

it follows that θ_4 is rejected at the same stage, contradicting the hypothesis Similarly if θ_4 is accepted initially at the mth stage, then

$$\sum_{\alpha=1}^{m} x_\alpha \geqq \frac{\sigma^2}{\Delta} \log A + a_2 m.$$

Since

$$\frac{\sigma^2}{\Delta} \log A + a_2 m > \frac{\sigma^2}{\Delta} \log A + a_1 m$$

it follows that θ_1 is rejected at the same or at an earlier stage, contradicting the assumption that the acceptance of θ_4 is an initial decision. Hence θ_1 and θ_4 cannot both be accepted.

A geometrical representation of the rule R is given in Figure 1.

R can now be described as follows: Continue taking observations until an acceptance region (shaded area) is reached or both dashed lines are crossed. In the former case, stop and accept as shown above. In the latter case stop and accept H_2.

The proof above that θ_1 and θ_4 cannot both be accepted consists of noting that a point below the acceptance line for θ_1 is already below the rejection line for θ_4 and that a point above the acceptance line for θ_4 is already above the rejection line for θ_1.

If $(\hat{A}, \hat{B}) \neq (A, B)$, a necessary and sufficient condition for the impossibility of accepting θ_1 and θ_4 is that at $n = 1$ the following inequalities should hold.

Rejection Number (of θ_1) for $R_1 \leqq$ Rejection Number (of θ_3) for R_2

and

Acceptance Number (of θ_1) for $R_1 \leqq$ Acceptance Number (of θ_3) for R_2.

In symbols

$$\frac{\sigma^2}{\Delta} \log A + a_1 \leqq \frac{\sigma^2}{\Delta} \log \hat{A} + a_2$$

and

$$\frac{\sigma^2}{\Delta} \log B + a_1 \leqq \frac{\sigma^2}{\Delta} \log \hat{B} + a_2.$$

These can be written as

$$\frac{A}{\hat{A}} \leqq e^{d\Delta/\sigma^2} \quad \text{and} \quad \frac{B}{\hat{B}} \leqq e^{d\Delta/\sigma}$$

respectively, where $d = a_2 - a_1$.

Since $\frac{d\Delta}{\sigma^2} > 0$, the above inequalities are certainly fulfilled when

$$\frac{B}{\bar{B}} \leqq 1 \quad \text{and} \quad \frac{A}{\bar{A}} \leqq 1. \tag{2.1}$$

In what follow in this paper, we shall restrict ourselves to cases where acceptance of both θ_1 and θ_4 is impossible, even if this is not stated explicitly.

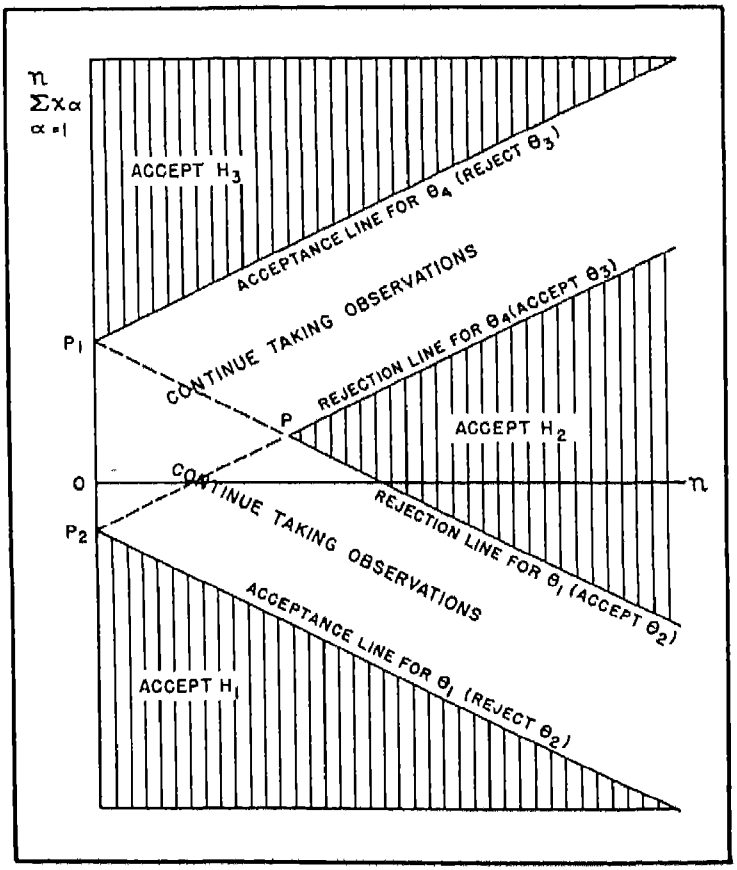

FIGURE 1

3. Derivation of OC Functions. Let $L(H_i \mid \theta, R)$ denote the probability of accepting H_i when θ is the true mean and R is the sequential rule used. Let H_{θ_i} denote the hypothesis that $\theta = \theta_i$. Since, as shown above, H_1 is accepted if and only if θ_1 is accepted, we have

$$L(H_1 \mid \theta, R) = L(H\theta_1 \mid \theta, R_1). \tag{3.1}$$

Similarly,

$$L(H_2 \mid \theta, R) = L(H_{\theta_4} \mid \theta, R_4). \tag{3.2}$$

From the fact that R_1 and R_2 each terminate at some finite stage with probability one, it follows that R will terminate at some finite stage with probability one. Hence

$$L(H_2 \mid \theta, R) = 1 - L(H_1 \mid \theta, R) - L(H_3 \mid \theta, R). \tag{3.3}$$

From pp. 50–52 of [1], the following equations are obtained.

$$L(H_1 \mid \theta, R) = L(H_{\theta_1} \mid \theta, R_1) = \frac{A^{h_1} - 1}{A^{h_1} - B^{h_1}} \tag{3.4}$$

where

$$h_1 = h_1(\theta) = \frac{\theta_2 + \theta_1 - 2\theta}{\theta_2 - \theta_1} = \frac{a_1 - \theta}{\frac{\Delta}{2}}$$

and

$$L(H_{\theta_3} \mid \theta, R_2) = \frac{\hat{A}^{h_2} - 1}{\hat{A}^{h_2} - \hat{B}^{h_2}} \tag{3.5}$$

where

$$h_2 = h_2(\theta) = \frac{\theta_4 + \theta_3 - 2\theta}{\theta_4 - \theta_3} = \frac{a_2 - \theta}{\frac{\Delta}{2}}.$$

These equations involve an approximation, as explained in [1]

Hence

$$L(H_3 \mid \theta, R) = L(H_{\theta_4} \mid \theta, R_2) = 1 - L(H_{\theta_3} \mid \theta, R_2) = \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}} \tag{3.6}$$

and

$$L(H_2 \mid \theta, R) = 1 - \frac{A^{h_1} - 1}{A^{h_1} - B^{h_1}} - \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}} = \frac{1 - B^{h_1}}{A^{h_1} - B^{h_1}} - \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}}. \tag{3.7}$$

Since $L(H_1 \mid \theta, R) = L(H_{\theta_1} \mid \theta, R_1)$, it follows that $L(H_1 \mid \theta, R)$ is a monotonically decreasing function of θ and that

$$L(H_1 \mid -\infty, R) = 1, \qquad L(H_1 \mid \infty, R) = 0$$
$$L(H_1 \mid \theta_1, R) = 1 - \alpha; \qquad L(H_1 \mid \theta_2, R) = \beta$$
$$L(H_1 \mid a_1, R) = \frac{\log A}{\log A + |\log B|}.$$

Similarly, since $L(H_3 \mid \theta, R) = 1 - L(H_{\theta_3} \mid \theta, R_2)$, it follows that $L(H_3 \mid \theta, R)$ is a monotonically increasing function of θ and that

$$L(H_3 \mid -\infty, R) = 0; \qquad L(H_3 \mid \infty, R) = 1$$
$$L(H_3 \mid \theta_3, R) = \hat{\alpha}; \qquad L(H_3 \mid \theta_4, R) = 1 - \hat{\beta}$$
$$L(H_3 \mid a_2, R) = \frac{|\log \hat{B}|}{\log \hat{A} + |\log \hat{B}|}.$$

Since $L(H_2 \mid \theta, R) = 1 - L(H_1 \mid \theta, R) - L(H_3 \mid \theta, R)$ it follows easily from the above results that

$$L(H_2 \mid -\infty, R) = 0; \qquad L(H_2 \mid \infty, R) = 0$$

$$L(H_2 \mid \theta, R) < \alpha \quad \text{for} \quad \theta < \theta_1; \qquad L(H_2 \mid \theta, R) < \hat{\beta} \quad \text{for} \quad \theta > \theta_4$$

$$\frac{|\log B|}{\log A + |\log B|} - \hat{\alpha} < L(H_2 \mid a_1, R) < \frac{|\log B|}{\log A + |\log B|}$$

$$\frac{\log \hat{A}}{\log \hat{A} + |\log \hat{B}|} - \beta < L(H_2 \mid a_2, R) < \frac{\log \hat{A}}{\log \hat{A} + |\log \hat{B}|}$$

$$1 - \beta - \hat{\alpha} < L(H_2 \mid \theta, R) < 1 \quad \text{for} \quad \theta_2 \leqq \theta \leqq \theta_3.$$

4. Probability of Correct Decision. Denote the probability of a correct decision by $L(\theta/R)$. It is defined as follows:

Interval	*Correct Decisions*	$L(\theta \mid R)$
$\theta \leqq \theta_1$	acceptance of H_1	$L(H_1 \mid \theta, R)$
$\theta_1 < \theta < \theta_2$	acceptance of H_1 or H_2	$L(H_1 \mid \theta, R) + L(H_2 \mid \theta, R)$
$\theta_2 \leqq \theta \leqq \theta_3$	acceptance of H_2	$L(H_2 \mid \theta, R)$
$\theta_3 < \theta < \theta_4$	acceptance of H_2 or H_3	$L(H_2 \mid \theta, R) + L(H_3 \mid \theta, R)$
$\theta_4 \leqq \theta$	acceptance of H_3	$L(H_3 \mid \theta, R)$

It should be noted that at points of discontinuity, $L(\theta, \mid R)$ is defined as the smaller of the two limiting values.

We shall now discuss some monotonicity properties of the function $L(\theta \mid R)$. From the fact that $L(H_{\theta_1} \mid \theta, R_1)$ and $L(H_{\theta_2} \mid \theta, R_2)$ are continuous with continuous first and second derivatives and are monotonically decreasing for all θ with a single point of inflection in the intervals $\theta_1 < \theta < \theta_2$ and $\theta_3 < \theta < \theta_4$ respectively, it follows that

(i) $L(\theta \mid R)$ is monotonically decreasing with negative curvature for $-\infty < \theta \leqq \theta_1$.

(ii) $L(\theta \mid R)$ is monotonically increasing with negative curvature for $\theta_4 \leqq \theta < \infty$.

Making use of (3.3) we have further

(iii) $L(\theta \mid R)$ is monotonically decreasing with negative curvature for $\theta_1 < \theta < \theta_2$.

(iv) $L(\theta \mid R)$ is monotonically increasing with negative curvature for $\theta_3 < \theta < \theta_4$.

(v) For $\theta_2 \leqq \theta \leqq \theta_3$, $\dfrac{d}{d\theta} L(\theta \mid R) = -\left[\dfrac{d}{d\theta} L(H_1 \mid \theta, R) + \dfrac{d}{d\theta} L(H_3 \mid \theta, R)\right]$ is decreasing, since $\dfrac{d}{d\theta} L(H_1 \mid \theta, R)$ and $\dfrac{d}{d\theta} L(H_3 \mid \theta, R)$ are increasing. In other words $L(\theta \mid R)$ has negative curvature for $\theta_2 \leqq \theta \leqq \theta_3$.

In the special case when $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and the origin is taken at $\frac{a_1 + a_2}{2}$ for the sake of convenience, it is easy to see that $L(\theta \mid R)$ is symmetric with respect to the origin and, because of (v), has a local maximum at $\theta = 0$.

5. Choice of the constants A, B, $\hat{A}$, $\hat{B}$ to insure prescribed Lower Bounds for $L(\theta \mid R)$. We shall deal here with the question of choosing A, B, $\hat{A}$ and $\hat{B}$ such that $L(\theta \mid R) \geqq 1 - \gamma_1$ when $\theta \leqq \theta_1$, $L(\theta \mid R) \geqq 1 - \gamma_2$ when $\theta_1 < \theta < \theta_4$, and $L(\theta \mid R) \geqq 1 - \gamma_3$ when $\theta \geqq \theta_4$. From the monotonic properties of the correct decision function it is only necessary to insure that

$$(5.1)\quad L(\theta_1 \mid R) = 1 - \gamma_1, L(\theta_2 \mid R) = L(\theta_3 \mid R) = 1 - \gamma_2 \text{ and } L(\theta_4 \mid R) = 1 - \gamma_3.$$

The following relations will be needed:

$$h_1(\theta_1) = h_2(\theta_3) = 1 = -h_1(\theta_2) = -h_2(\theta_4)$$

$$h_2(\theta_2) = \frac{\theta_3 + \theta_4 - 2\theta_2}{\Delta} = \frac{d - \frac{\Delta}{2}}{\frac{\Delta}{2}} = r \quad \text{(say)}$$

$$h_1(\theta_3) = \frac{\theta_1 + \theta_2 - 2\theta_3}{\Delta} = \frac{-d + \frac{\Delta}{2}}{\frac{\Delta}{2}} = -r$$

where $d = \theta_4 - \theta_2 = \theta_3 - \theta_1 = a_2 - a_1$.

The following four equations are obtained from (5.1):

$$(5.2)\quad 1 - L(H_1 \mid \theta_1, R) = L(H_{\theta_2} \mid \theta_1, R_1) = \frac{1 - B}{A - B} = \gamma_1$$

$$1 - L(H_2 \mid \theta_2, R) = L(H_1 \mid \theta_2, R) + L(H_3 \mid \theta_2, R)$$

$$(5.3)\qquad = \frac{B(A - 1)}{A - B} + \left[\frac{1 - \hat{B}^r}{\hat{A}^r - \hat{B}^r}\right] = \gamma_2$$

$$1 - L(H_2 \mid \theta_3, R) = L(H_3 \mid \theta_3, R) + L(H_1 \mid \theta_3, R)$$

$$(5.4)\qquad = \frac{1 - \hat{B}}{\hat{A} - \hat{B}} + \left[\frac{B^r(A^r - 1)}{A^r - B^r}\right] = \gamma_2$$

$$(5.5)\quad 1 - L(H_3 \mid \theta_4, R) = L(H_{\theta_3} \mid \theta_4, R_2) = \frac{\hat{B}(\hat{A} - 1)}{\hat{A} - \hat{B}} = \gamma_3.$$

The "bracketed terms" represent quantities less than $\hat{\alpha}$ and β respectively and if r is sufficiently large they can be neglected. This will be made more precise but first let us note the results of neglecting the bracketed terms.

From (5.2) and (5.3) we obtain

$$(5.6)\qquad B(1 - \gamma_1) = \gamma_2, \quad \text{whence} \quad B = \frac{\gamma_2}{1 - \gamma_1}.$$

From (5.2) and (5.6)

$$A = \frac{1 - B(1 - \gamma_1)}{\gamma_1} \quad \text{whence} \quad A = \frac{1 - \gamma_2}{\gamma_1}. \tag{5.7}$$

Since the last two equations are obtained from the first two by the permutation $A \to \hat{A}, B \to \hat{B}, \gamma_1 \to \gamma_2, \gamma_2 \to \gamma_3$, we have

$$\hat{B} = \frac{\gamma_3}{1 - \gamma_2}$$

$$\hat{A} = \frac{1 - \gamma_3}{\gamma_2}.$$

If $\gamma_1 = \gamma_2 = \gamma_3 = \gamma$ (say) then $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}} = \frac{1 - \gamma}{\gamma}$.

We shall consider the bracketed quantities negligible if the result of neglecting them produces a change of less than 20% in $[1 - L(\theta \mid R)]$ at $\theta = \theta_2, \theta_3$ respectively, i.e., if

$$\frac{1 - \hat{B}^r}{\hat{A}^r - {}^r\hat{B}} = \frac{1 - \left(\frac{\gamma_3}{1 - \gamma_2}\right)^r}{\left(\frac{1 - \gamma_3}{\gamma_2}\right)^r - \left(\frac{\gamma_3}{1 - \gamma_2}\right)^r} \leqq \frac{\gamma_2}{5} \tag{5.8}$$

and

$$\frac{B^r(A^r - 1)}{A^r - B^r} = \frac{\left(\frac{\gamma_2}{1 - \gamma_1}\right)^r \left[\left(\frac{1 - \gamma_2}{\gamma_1}\right)^r - 1\right]}{\left(\frac{1 - \gamma_2}{\gamma_1}\right)^r - \left(\frac{\gamma_2}{1 - \gamma_1}\right)^r} \leqq \frac{\gamma_2}{5}. \tag{5.9}$$

Inequality, (5.9) can be written as

$$\frac{\gamma_2^r[(1 - \gamma_2)^r - \gamma_1^r]}{(1 - \gamma_2)^r(1 - \gamma_1)^r - \gamma_1^r\gamma_2^r} \leqq \frac{\gamma_2}{5}$$

or

$$(1 - \gamma_2)^r\left[\gamma_2^r - \frac{\gamma_2}{5}(1 - \gamma_1)^r\right] \leqq (\gamma_1\gamma_2)^r\left(1 - \frac{\gamma_2}{5}\right).$$

This will certainly hold if

$$\gamma_2^r \leqq \frac{\gamma_2}{5}(1 - \gamma_1)^r$$

or if

$$\left(\frac{\gamma_2}{1 - \gamma_1}\right)^r \leqq \frac{\gamma_2}{5}.$$

Assume that γ_1, γ_2 and γ_3 are each less than $\frac{1}{2}$. Then the last inequality can be

written as

$$r \geqq \frac{\log\left(\frac{5}{\gamma_2}\right)}{\log\left(\frac{1-\gamma_1}{\gamma_2}\right)}. \tag{5.10}$$

Starting with (5 8) the same relation is obtained except that γ_1 is replaced by γ_3, namely

$$r \geqq \frac{\log \frac{5}{\gamma_2}}{\log \frac{1-\gamma_3}{\gamma_2}}. \tag{5.11}$$

Let

$$k = \frac{\log \frac{5}{\gamma_2}}{\log \frac{1-\bar{\gamma}}{\gamma_2}}$$

where $\bar{\gamma}$ is the larger of γ_1 and γ_3. Then k is the larger of the right hand members of (5.10) and (5 11). Then for (5 8) and (5.9) to hold it is sufficient that

$$r \geqq k.$$

If $\gamma_2 = .05$ and $0 < \gamma_1, \gamma_3 < .1$ then k is approximately $\frac{2}{1.3} = 1\,54$. If $\gamma_2 = 01$ and $0 < \gamma_1, \gamma_3 < .1$ then k is approximately $\frac{2.7}{2} = 1.35$.

We shall now investigate under what conditions the approximate solution obtained above for A, B, $\hat{A}$, $\hat{B}$ are such that acceptance of both θ_1 and θ_4 is impossible.

It follows from (2 1) that the following pair of inequalities are sufficient for the impossibility of accepting both θ_1 and θ_4 :

$$\frac{A}{\hat{A}} = \frac{\gamma_2}{\gamma_1}\frac{1-\gamma_2}{1-\gamma_3} \leqq 1; \qquad \frac{B}{\hat{B}} = \frac{\gamma_2}{\gamma_3}\frac{1-\gamma_2}{1-\gamma_1} \leqq 1. \tag{5.12}$$

If $\gamma_1 \neq \gamma_3$ let the smaller and larger of the pair (γ_1, γ_3) be denoted by $\underline{\gamma}$ and $\bar{\gamma}$ respectively. Since $1 - \underline{\gamma} > 1 - \bar{\gamma}$, then

$$\frac{\gamma_2(1-\gamma_2)}{\bar{\gamma}(1-\underline{\gamma})} < \frac{\gamma_2(1-\gamma_2)}{\underline{\gamma}(1-\bar{\gamma})}$$

and we need only consider one of the two inequalities in (5.12). The condition $\gamma_2 < \underline{\gamma}$ will in general satisfy (5.12). More precisely if all the γ's are restricted to the interval (0, .1) then

$$\frac{9}{10} \leqq \frac{1-\gamma_2}{1-\underline{\gamma}} < \frac{1-\gamma_2}{1-\bar{\gamma}} \leqq \frac{10}{9}$$

and it is sufficient for the validity of (5.12) that $\gamma_2 \leqq (.9)\,\underline{\gamma}$.

If $\gamma_1 = \gamma_3 = \gamma$ (say) then the two inequalities reduce to one

$$\gamma_2^2 - \gamma_2 + \gamma - \gamma^2 \geqq 0$$

which can be written as

$$(\gamma_2 - \gamma)(\gamma_2 - 1 + \gamma) \geqq 0.$$

Since the inequality $\gamma_2 \geqq 1 - \gamma$ is impossible when all γ's are $< \frac{1}{2}$, we see that $\gamma_2 \leq \gamma$ is sufficient for the validity of (5.12) when $\gamma_1 = \gamma_3 = \gamma < \frac{1}{2}$.

There remains the problem of finding an approximate solution for equations (5.2) to (5.5) when $r < k$. Since

$$r = \frac{d - \frac{\Delta}{2}}{\frac{\Delta}{2}} = \frac{\theta_3 - \theta_2 + \frac{\Delta}{2}}{\frac{\Delta}{2}} \geqq 1$$

we merely have to consider the interval $1 \leqq r < k$.

The following approximations are used

$$\begin{aligned} &\frac{1 - B}{A - B} \sim \frac{1}{A}; \qquad \frac{B(A-1)}{A - B} \sim B; \qquad \frac{1 - \hat{B}^r}{\hat{A}^r - \hat{B}^r} \sim \frac{1}{\hat{A}^r} \\ &\frac{1 - \hat{B}}{\hat{A} - \hat{B}} \sim \frac{1}{\hat{A}}; \qquad \frac{B^r(A^r - 1)}{A^r - B^r} \sim B^r; \qquad \frac{\hat{B}(\hat{A} - 1)}{\hat{A} - \hat{B}} \sim \hat{B}, \end{aligned} \tag{5.13}$$

which upon substitution yield

$$A = \frac{1}{\gamma_1} \tag{5.14}$$

$$\hat{B} = \gamma_3 \tag{5.15}$$

$$B + \frac{1}{\hat{A}^r} = \gamma_2 \tag{5.16}$$

$$\frac{1}{\hat{A}} + B^r = \gamma_2. \tag{5.17}$$

Subtraction of (5.17) from (5.16) shows that $B = \frac{1}{\hat{A}}$ is a solution. Substituting this result back in (5.16) leads to the equation

$$B + B^r = \gamma_2. \tag{5.18}$$

It can easily be verified that between zero and unity this equation has exactly one root. Since $1 \leqq r < \infty$, the root of the above equation lies between $\frac{\gamma_2}{2}$ and γ_2.

Taking γ_2 as a first approximation for B and substituting $\gamma_2 + \epsilon$ for B in (5.18), we obtain

$$\epsilon + (\gamma_2 + \epsilon)^r = 0.$$

Expanding $(\gamma_2 + \epsilon)^r$ in a power series in ϵ and neglecting second and higher order terms, the above equation gives

$$\epsilon \sim \frac{\gamma_2^r}{1 + r\gamma_2^{r-1}}.$$

Thus,

$$B = \frac{1}{A} \sim \gamma_2 - \frac{\gamma_2^r}{1 + r\gamma_2^{r-1}} = \frac{\gamma_2[1 + (r-1)\gamma_2^{r-1}]}{1 + r\gamma_2^{r-1}}. \tag{5.19}$$

It is necessary to investigate under what conditions the above approximate solution satisfies (5.2) to (5.5) to within a 20% error in $[1 - L(\theta/R)]$, i.e., such that

$$-\frac{\gamma_1}{5} < \frac{\gamma_1(1-B)}{1-\gamma_1 B} - \gamma_1 < \frac{\gamma_1}{5} \tag{5.20}$$

$$-\frac{\gamma_3}{5} < \frac{\gamma_3(1-B)}{1-\gamma_3 B} - \gamma_3 < \frac{\gamma_3}{5} \tag{5.21}$$

$$-\frac{\gamma_2}{5} < \frac{B(1-\gamma_1)}{1-\gamma_1 B} + \frac{B^r(1-\gamma_3^r)}{1-(\gamma_3 B)^r} - \gamma_2 < \frac{\gamma_2}{5} \tag{5.22}$$

$$-\frac{\gamma_2}{5} < \frac{B(1-\gamma_3)}{1-\gamma_3 B} + \frac{B^r(1-\gamma_1^r)}{1-(\gamma_1 B)^r} - \gamma_2 < \frac{\gamma_2}{5} \tag{5.23}$$

where for B the value in (5.19) is understood.

It can be shown that if γ_1, γ_2, γ_3, are each between zero and .1 then the inequalities (5.20) to (5.23) hold. Furthermore it can be shown that if, in addition $\gamma_2 \leqq \min(\gamma_1, \gamma_3)$ then also the inequalities (2.1) hold. The latter inequalities are sufficient to ensure the impossibility of accepting both θ_1 and θ_4.

6. Bounds for the ASN Function. First we shall derive lower bounds for the ASN function. Let $E(n/\theta, R)$ denote the expected value of n when θ is the true mean and R is the sequential rule employed. For $\theta < \theta_2$ the probability of coming to a decision first with R_2 is large and therefore

$$E(n/\theta, R) \sim E(n/\theta, R_1) \quad \theta < \theta_2.$$

From the definition of R it follows that

$$E(n/\theta, R) > E(n/\theta, R_1) \qquad \text{for all } \theta.$$

Hence $E(n/\theta, R_1)$ serves as a close lower bound when $\theta < \theta_2$.

Similarly

$$E(n/\theta, R) \sim E(n/\theta, R_2) \qquad \text{for } \theta > \theta_3$$

$$E(n/\theta, R) > E(n/\theta, R_2) \qquad \text{for all } \theta$$

Hence $E(n/\theta, R_2)$ serves as a close lower bound for $\theta > \theta_3$.

Combining the above we have

$$E(n/\theta, R) > \text{Max}\,[E(n/\theta, R_1),\ E(n/\theta, R_2)] \tag{6.1}$$

where, neglecting the excess over the boundary,

$$E(n/\theta, R_1) = \frac{L(H_{\theta_1}/\theta, R_1) \log B + L(H_{\theta_2}/\theta, R_1) \log A}{\frac{\Delta}{\sigma^2}(\theta - a_1)} \tag{6.2}$$

$$E(n/\theta, R_2) = \frac{L(H_{\theta_3}/\theta, R_2) \log \hat{B} + L(H_{\theta_4}/\theta, R_2) \log \hat{A}}{\frac{\Delta}{\sigma_2}(\theta - a_2)} \tag{6.3}$$

Formula (6.1) gives a valid lower bound over the whole range of θ, but this lower bound will not be very close in the interval (θ_2, θ_3), particularly in the neighbourhood of the mid-point $\frac{\theta_2 + \theta_3}{2}$. The authors were not able to find any simple method for obtaining a closer lower bound in this interval. The upper bound given later in this section will, however, be fairly close also in the interval (θ_2, θ_3) and can be used as an approximation to the exact value.

We shall now derive upper bounds for the ASN function. Let R_1^* be the following rule: "Continue to take observations until R_1 accepts θ_1." Since this implies the rejection of θ_4 at the same or at a previous stage, it follows that R must terminate not later than R_1^*. Hence

$$E(n/\theta, R_1^*) \geqq E(n/\theta, R). \tag{6.4}$$

As a matter of fact one can easily verify that $E(n/\theta, R_1^*) > E(n/\theta, R)$. Thus $E(n/\theta, R_1^*)$ is an upper bound for $E(n/\theta, R)$. This upper bound will be close when the probability of accepting θ_1 is high, i.e., for $\theta \leqq \theta_1$.

By the general formula

$$E(n) = \frac{E\left(\sum_{i=1}^{n} z_i\right)}{E(z)}$$

(see p. 53 [1]) we obtain, upon neglecting the excess over the boundary,

$$E(n/\theta, R_1^*) = \frac{\log B}{\frac{\Delta}{\sigma^2}(\theta - a_1)}. \tag{6.5}$$

This coincides with (6.2) when $L(H_{\theta_2}/\theta, R_1) = 0$.

Similarly, if R_2^* denotes the rule of continuing until R_2 accepts θ_4, then

$$E(n/\theta, R_2^*) > E(n/\theta, R) \tag{6.6}$$

$$E(n/\theta, R_2^*) = \frac{\log \hat{A}}{\frac{\Delta}{\sigma^2}(\theta - a_2)} \tag{6.7}$$

and this will be a close upper bound for $\theta \geqq \theta_4$.

If $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and if $a_1 + a_2 = 0$ the above results reduce to

$$E(n/\theta, R) \lesssim E(n/\theta, R_1^*) = \frac{-h}{\lambda + \theta} \qquad \text{for} \qquad \theta \leqq \theta_1 \tag{6.8}$$

$$E(n/\theta, R) \lesssim E(n/\theta, R_2^*) = \frac{h}{\theta - \lambda} \qquad \text{for} \qquad \theta \geqq \theta_4 \tag{6.9}$$

where the symbol $\lesssim$ stands for a close inequality, and where

$$h = \frac{\sigma^2}{\Delta} \log A \qquad \text{and} \qquad \lambda = a_2 = -a_1 .$$

To establish an upper bound for the ASN function in the interval $\theta_2 < \theta < \theta_3$ we shall restrict ourself to the case where $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$. These relations are fulfilled by the approximate values of A, B, $\hat{A}$, $\hat{B}$ suggested in section 5 when $\gamma_1 = \gamma_2 = \gamma_3$ and $r \geqq k$. We shall choose the origin to be at $\frac{a_1 + a_2}{2}$, i.e., we put $\frac{a_1 + a_2}{2} = 0$. Then the vertex P of the triangle (P_1, P_2, P) in diagram 1 lies on the abscissa axis and $OP_1 = OP_2 = h$ The abscissa of the vertex P is $\frac{h}{\lambda} = N$ (say) where $\lambda = a_2 = -a_1$ Let $y = \sum_{i=1}^{N} X_i$ represent the sum of the first N observations Let R_{23} denote the rule: "Continue until both θ_2 and θ_3 are accepted". This is tantamount to neglecting the two outer lines in diagram 1, i.e , the acceptance lines for θ_1 and θ_4. Then clearly,

$$E(n/\theta, R_{23}) > E(n/\theta, R). \tag{6 10}$$

When θ lies between θ_2 and θ_3 this inequality will be close, since the probability of crossing either of the two outer lines is then small

However $E(n/\theta, R_{23})$ was found difficult to compute and it was necessary to consider instead the rule R'_{23} : "Take N observations. If $y = \sum_{i=1}^{N} X_i < 0$ then continue until θ_2 is accepted. If $y > 0$ then continue until θ_3 is accepted".[3] Clearly,

$$E(n/\theta, R'_{23}) > E(n/\theta, R_{23}). \tag{6.11}$$

This inequality, however, will be close only if the probability of concluding the test before N observations, given that $\theta_2 < \theta < \theta_3$, is small.

Some investigations by the authors seem to indicate that the inequality (6 11) will be close when $\Delta < \lambda$. This inequality is likely to be fulfilled in practical problems

We shall now proceed to determine the value of $E(n/\theta, R'_{23})$. Neglecting the excess over the boundary, we have

$$E\left(n/\theta, R'_{23}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} + \frac{y}{\lambda - \theta} \qquad \text{for} \qquad y > 0 \tag{6.12}$$

[3] The event $y = 0$ has probability zero and it is indifferent what rule is adopted for that case.

and

$$(6.13) \qquad E\left(n/\theta, R'_{23}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} - \frac{y}{\lambda + \theta} \qquad \text{for} \qquad y < 0$$

where, for any condition C, $E(n/\theta, R, C)$ denotes the conditional expected value of n given that the true mean is θ, that R is the sequential rule used and that the condition C is fulfilled.

Multiplying with the density of y and then integrating with respect to y, we obtain after simplification

$$(6.14) \quad E(n/\theta, R'_{23}) = \frac{1}{\lambda^2 - \theta^2}\left[h\lambda + 2h\theta\phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right) + 2\sigma\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\theta^2/2\lambda\sigma^2)}\right]$$

where $\phi(x) = \int_0^x \frac{e^{-(y^2/2)}}{\sqrt{2\pi}}\, dy$, and $\theta_2 < \theta < \theta_3$.

In particular, for $\theta = 0$ we get

$$(6.15) \qquad E(n/\theta = 0, R'_{23}) = \frac{h}{\lambda} + \frac{\sigma}{\lambda^2}\sqrt{\frac{2h\lambda}{\pi}}.$$

To establish a close upper bound for $\theta_3 < \theta < \theta_4$ we must bring the line of acceptance of θ_4 into account. The line of acceptance of θ_1 can be disregarded since the probability of accepting θ_1 is very small.

We therefore define the rule R_{34} as follows:

"Continue with R_1 until θ_2 is accepted and with R_2 until either θ_3 or θ_4 is accepted."

Since the ASN function for R_{34} is difficult to compute we define a modified rule R'_{34} as follows:

"Proceed to take $N\left(=\frac{h}{\lambda}\right)$ observations without regard to any rule. If $y = \sum_{i=1}^{N} X_i < 0$ then continue only with R_1 until θ_2 is accepted. If $0 < y < 2h$ then continue only with R_2 until either θ_3 or θ_4 is accepted. If $y \geqq 2h$ then stop taking observations and accept H_3."

It is clear that the following inequalities hold

$$(6.16) \qquad E(n/\theta, R'_{34}) > E(n/\theta, R_{34}) > E(n/\theta, R).$$

The proximity of $E(n/\theta, R_{34})$ and $E(n/\theta, R)$, as stated above, is based on the fact that the probability of accepting θ_1, when $\theta_3 < \theta < \theta_4$, is small.

The proximity of $E(n/\theta, R_{34})$ and $E(n/\theta, R'_{34})$ is assured if the probability of terminating with R_{34} (and with R) before N observations is small. It can be shown that the latter condition is fulfilled when $\Delta < \lambda$. In terms of the quantity r defined in Section 5 this can be written as $r > 3$.

To determine the value of $E(n/\theta, R'_{34})$ the following two preliminary results will be needed:

If $0 < y < 2h$,

$$(6.17)\quad E\left(n/\theta, R'_{34}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} + \frac{2h - y - 2h\left[\dfrac{1 - e^{-(2/\sigma^2)(\lambda-\theta)(2h-y)}}{1 - e^{-(4h/\sigma^2)(\lambda-\theta)}}\right]}{\theta - \lambda} = C \text{ (say)}.$$

If $y < 0$,

$$(6.18)\quad E\left(n/\theta, R'_{34}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} - \frac{y}{\lambda + \theta} = D \text{ (say)}.$$

Both are easily obtained from formula (7.25) on p. 123 of [1].
Multiplying with the density of y and integrating with respect to y, we obtain after simplification

$$(6.19)\quad \begin{aligned} E(n/\theta, R'_{34}) = \frac{h}{\lambda} &+ \left[\phi\left(\frac{2\lambda - \theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right) + \phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right)\right] \\ &\cdot \frac{h}{(\lambda - \theta)}\left(\frac{\theta}{\lambda} - \frac{2e^{-(2h(\lambda-\theta)/\sigma^2)}}{1 + e^{-2h(\lambda-\theta)/\sigma^2}}\right) \\ &+ \frac{\sigma}{\lambda(\lambda - \theta)}\sqrt{\frac{h\lambda}{2\pi}}\left[e^{-(h\theta^2/2\lambda\sigma^2)} - e^{-h(2\lambda-\theta)^2/2\lambda\sigma^2}\right] \\ &- \frac{h\theta}{2\lambda(\lambda + \theta)}\left[1 - 2\phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right)\right] + \frac{\sigma}{\lambda(\lambda + \theta)}\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\theta^2/2\lambda\sigma^2)}. \end{aligned}$$

Formula (6.19) is an improvement on (6.14) as it will give for any θ a smaller upper bound, but in the neighborhood of the origin the difference is insignificant.

For $\theta = \lambda$ we obtain from (6.19) using L'Hopital's rule

$$(6.20)\quad \begin{aligned} E(n/\lambda, R'_{34}) = \frac{h^2}{\sigma^2} &- \frac{h}{4\lambda\sigma^2}(4h\lambda - 3\sigma^2) \\ &\cdot\left[1 - 2\phi\left(\frac{\sqrt{h\lambda}}{\sigma}\right)\right] + \left(\frac{h\lambda + \sigma^2}{2\lambda^2\sigma}\right)\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\lambda/2\sigma^2)}. \end{aligned}$$

If $\dfrac{\sqrt{h\lambda}}{\sigma} > 2.5$, the above formula can be approximated by

$$(6.21)\quad E(n/\lambda, R'_{34}) \sim \frac{h^2}{\sigma^2} + \frac{2\sigma}{\lambda^2}\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\lambda/2\sigma^2)}.$$

Since the right hand member above lies between $\dfrac{h^2}{\sigma^2}$ and $(1.002)\dfrac{h^2}{\sigma^2}$ when $\dfrac{\sqrt{h\lambda}}{\sigma} > 2.5$ then for practical purposes

$$(6.22)\quad E(n/\lambda, R'_{34}) \sim \frac{h^2}{\sigma^2} \qquad \text{when } \left(\frac{\sqrt{h\lambda}}{\sigma} > 2.5\right).$$

An upper bound for $E(n/\theta, R)$ for $\theta_1 < \theta < \theta_4$ can be obtained by defining R_{12} and R'_{12} in an analagous way to R_{34} and R'_{34} Because of reasons of symmetry, $E(n/\theta, R'_{12})$ can be obtained from (6.19) by replacing θ by $-\theta$.

The method used for obtaining upper bounds for $E(n/\theta, R)$ can easily be extended to the more general case when the equalities $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ do not necessarily hold. However, the resulting formulas are more cumbersome and we shall merely give without proof the upper bound corresponding to (6.14). This upper bound becomes

$$E(n/\theta, R'_{23}) = N + \left(\frac{N\theta - h_3}{\lambda - \theta}\right)\left[\frac{1}{2} - \phi(a)\right] + \left(\frac{h_3 - N\theta}{\lambda + \theta}\right)\left[\frac{1}{2} - \phi(b)\right] + \sigma\sqrt{\frac{N}{2\pi}}\left[\frac{e^{-a^2/2}}{\lambda - \theta} + \frac{e^{-b^2/2}}{\lambda + \theta}\right]$$

where

$$h_{11} = \frac{\sigma^2}{\Delta}\log A \qquad h_{10} = \frac{\sigma^2}{\Delta}\log B$$

$$h_{21} = \frac{\sigma^2}{\Delta}\log \hat{A} \qquad h_{20} = \frac{\sigma^2}{\Delta}\log \hat{B}$$

$$a_2 = -a_1 = \lambda$$

$$N = \frac{h_{11} - h_{20}}{2\lambda}; \qquad a = \frac{h_3 - N\theta}{\sigma\sqrt{N}}; \qquad b = \frac{h_3 + N\theta}{\sigma\sqrt{N}}; \qquad h_3 = \frac{h_{11} + h_{20}}{2}.$$

7. An Example. We shall consider the following example

$\sigma^2 = 1$, $\theta_1 = -\frac{5}{16}$, $\theta_2 = -\frac{3}{16}$, $\theta_3 = \frac{3}{16}$, $\theta_4 = \frac{5}{16}$, $\gamma_1 = \gamma_2 = \gamma_3 = \gamma = .029$

then

$$A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}} = \frac{1-\gamma}{\gamma} = 33.5 \qquad r = 7 \gg 3 > k \sim 1.47$$

and

$$h = \frac{\sigma^2}{\Delta}\log A = 28, \lambda = \frac{\theta_3 + \theta_4}{2} = \tfrac{1}{4}, \Delta = \theta_2 - \theta_1 = \theta_4 - \theta_3 = \tfrac{1}{8}.$$

Using formulas (6.1) and (6.7) the following upper and lower bounds were obtained

θ	$\frac{5}{16}$	$\frac{6}{16}$	$\frac{7}{16}$	$\frac{8}{16}$	$\frac{9}{16}$	$\frac{10}{16}$	$\frac{12}{16}$	$\frac{14}{16}$	$\frac{16}{16}$	$\frac{18}{16}$	$\frac{20}{16}$
Upper bound	448	224	149	112	89.6	74.7	56	44.8	37 3	32	28
Lower bound	421	224	149	112	89.6	74.7	56	44.8	37.3	32	28

Formulas (6.14) and (6.1) yield

θ	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$
Upper Bound	146	163	229	450
Lower Bound . .	112	149	224	421

In the neighborhood of the origin the true value is very nearly the upper bound. From formulas (6.19), (6 22) and (6.1) we obtain

θ	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{5}{16}$
Upper Bound . .	422	784 5	423
Lower Bound	421	784	421

As shown above for the end points of the indifference zone, (6.19) gives better results than (6.14) or (6.7). This is as it should be since (6.19) takes into account possibilities omitted in (6.14) and (6.7). The greater accuracy of (6.19) is offset by a slight increase in computation.

In the graph of the Bounds of the ASN function shown in Figure 2, a single curve is shown wherever the upper and lower bound are sufficiently close to each other.

Since (6 14) contains an even function of θ and since elsewhere the corresponding bounds are mirror images with respect to $\theta = 0$, the bounds for negative θ are exactly the same as those for the corresponding positive θ

Consider the following non-sequential rule applied to our problem. With a fixed number N_0 of observations compute the mean $\bar{x}$ and accept H_1 if $\bar{x}$ falls in the interval $(-\infty, a_1)$, accept H_2 if $\bar{x}$ falls in $[a_1, a_2]$ and accept H_3 if $\bar{x}$ falls in (a_2, ∞). This is certainly a reasonable procedure. One can also verify that no other non-sequential rule exists that is uniformly better (for all possible values of θ) than the one under consideration.

The two decision procedures become comparable if we introduce the indifference zones and define a wrong decision in the non-sequential case exactly as was done for our sequential procedure (see Section 1).

For the non-sequential case (just as in the sequential case) the probability of a wrong decision will be discontinuous at θ_1, θ_2, θ_3 and θ_4. At each of these points there will be a left-sided and right-sided limit, different from each other. As in the sequential case we shall take the probability of a wrong decision at a discontinuity point to be equal to the larger of the left and right hand limits One can easily verify that the maximum probability of a wrong decision occurs at $\theta = \theta_3$ (which is equal to the value at $\theta = \theta_2$).

We then determine N_0 by setting the maximum probability of a wrong decision equal to γ, i.e.

$$\phi\left(\frac{d - \Delta/2}{\sigma}\sqrt{N_0}\right) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma. \tag{7.1}$$

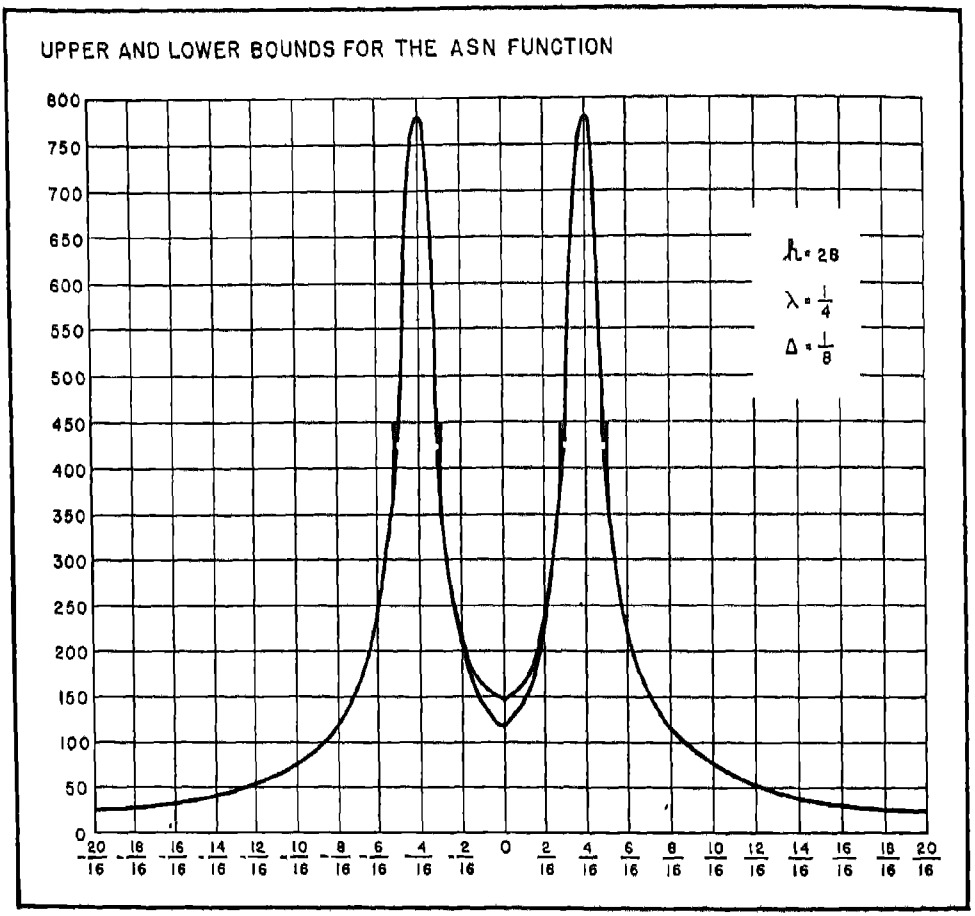

FIGURE 2

For the particular problem considered above, this gives $N_0 = 915.4$. Hence 916 observations are required in order to ensure that this non-sequential procedure will have the maximum probability $\gamma = .029$ of a wrong decision. This is to be compared with the maximum over all θ of the ASN function in the sequential procedure, which was 784.5.

Returning to (7.1) we shall derive lower and upper bounds for the root of that equation. Since

$$\infty > \frac{d - \Delta/2}{\sigma}\sqrt{N_0} \geqq \frac{\Delta}{2\sigma}\sqrt{N_0}$$

it is clear that the root of the equation

$$\phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma$$

is an upper bound for the root of (7.1) and that the root of the equation

$$\phi(\infty) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma$$

or

$$\phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = \frac{1}{2} - \gamma$$

is a lower bound for the root of (7.1). We shall compare the value of $x = \frac{\Delta}{2\sigma}\sqrt{N_0}$ with the value of $y = \frac{\Delta}{2\sigma}\sqrt{\operatorname{Max}_\theta \text{ ASN}}$. Since

$$\operatorname{Max}_\theta \text{ (ASN function)} \sim \frac{h^2}{\sigma^2} = \frac{\sigma^2}{\Delta^2}\left(\log\frac{1-\gamma}{\gamma}\right)^2 \text{ (for sufficiently small } \frac{\Delta}{d}\text{)}.$$

then

$$y = \frac{\Delta}{2\sigma}\sqrt{\operatorname{Max}_\theta \text{ ASN}} \sim \frac{1}{2}\log\frac{1-\gamma}{\gamma} \text{ (for sufficiently small } \frac{\Delta}{d}\text{)}.$$

The following table gives upper and lower bounds for x and the corresponding value of y for the type of example under consideration, i.e., when $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and $r \geqq k$.

γ	.001	.002	.005	.008	.01	.05	.1
$\underline{x}$ and $\bar{x}$	3.08–3.31	2.87–3.10	2.57–2.81	2.41–2.65	2.33–2.58	1.64–1.96	1.28–1.65
y	3.45	3.11	2.65	2.41	2.30	1.47	1.10

As the table shows[4] for $.1 > \gamma > .008$

$$x > \underline{x} > y$$

[4] Actually, the inequality in question is shown only for the values of γ given in the table. However it can be verified that the inequality remains valid for all values of γ between .1 and .008.

and hence

$$N_0 > \max_{\theta} \text{ASN} \qquad \text{(for sufficiently small } \frac{\Delta}{d}\text{)}.$$

The statement and the table above are not meant to delimit the region in which the sequential rule is superior to the non-sequential procedure.

REFERENCES

[1] ABRAHAM WALD, *Sequential Analysis*, John Wiley and Sons, 1947.

[2] P. ARMITAGE, "Some sequential tests of student's hypothesis," *Supplement to the Journal of the Royal Statistical Society*, *Vol.* 9 (1947) No. 2, p. 250.

MOMENTS OF RANDOM GROUP SIZE DISTRIBUTIONS[1]

By John W. Tukey

Princeton University

1. Summary. A number of practical problems involve the solution of a mathematical problem of the class described in the classical language of probability theory as follows. "A number of balls are *independently* distributed among a number of boxes, how many boxes contain no balls, 1 ball, 2 balls, 3 balls, and so on." Problems arising in the oxidation of rubber and the genetics of bacteria are discussed as applications.

A method is given of solving problems of this sort when "how many" is adequately answered by the calculation of means, variances, covariances, third moments, etc. The method is applied to a number of the simplest cases, where the number of balls is fixed, binomially distributed or Poisson and where the "sizes" of the boxes are equal or unequal.

2. Introduction. The distribution of the number of empty boxes has been investigated by Romanovsky in 1934 [3], and, apparently independently, by Stevens in 1937 [4]. Romanovsky investigated the case of N equal boxes and m balls for (i) the case where the balls are independent, and (ii) the case where there is a limit to the size of each box. He gives no motivation for the problem, and shows that certain limiting distributions approach normality. Stevens investigated the case of m independent balls for N boxes (i) of equal size, and (ii) of unequal size, and developed a useful approximation for the last case. Stevens was concerned with this problem in order to test box counts for nonrandomness by comparing the number of empty boxes with expectation. The reader interested in that problem is referred to his paper.

The results derived in Part II are based on the use of a chance generating function, a technique which applies easily to the case where the balls are independent. Thus Romanovsky's results for the case of boxes of limited size are neither included or extended For the other cases where the number of empty boxes has been considered, the results below seem to provide simple moments and cross-moments for the numbers of boxes with any number of balls to the extent previously available for the number of empty boxes. Both Romanovsky and Stevens investigated the actual distribution of the number of empty boxes. A similar investigation of the distribution of the number of b-ball boxes has *not* been carried out here.

3. A chemical problem. In studying the oxidation of rubber, Tobolsky and coworkers were led to propose the following problem: "If a mass of rubber originally consisted of N chains of equal length, if each chain can be broken at a

[1] Prepared in connection with research sponsored by the Office of Naval Research.

large number of places by the reaction with one oxygen molecule, if there are m oxygen molecules each equally likely to react at each link, and if mNp molecules have reacted, what is the probable number of original chains which are now in $b + 1$ parts as a result of b oxygen molecules having reacted with b of their links?

Here an original chain plays the role of a box and an oxygen molecule the role of a ball. The sort of numbers which may be taken as characteristic are:

$$N = 10^{18} \quad \text{(number of chains)},$$
$$m = 10^{16} \text{ to } 10^{20} \quad \text{(number of oxygen molecules)},$$
$$mp = 0.01 \text{ to } 100 \quad \text{(average breaks/chain)}.$$

Thus it is almost certainly going to be appropriate to use the results obtained by assuming N and m very large and $p = 1/N$ very small. We shall return to this example after discussing the general results.

4. A bacteriological problem. The experiments of Newcombe [1] on the irradiation and mutation of bacteria have prompted Pittendrigh to propose the following problem: "Suppose a large number of bacteria each contain m enzyme particles, which have been formed by the action of a nuclear gene. Suppose that irradiation destroys the nuclear gene in a certain fraction of the bacteria. Suppose three generations to occur, during which the m original enzyme particles are randomly distributed among the 8 descendants of an original bacterium. If a bacterium without either nuclear gene or enzyme particle is a recognizable mutant, what is the expected distribution of "families" with 0, 1, 2, 3, $\cdots$, 8 mutants?"

Here the enzyme particles are the balls, and the 8 descendants are the N boxes. We are interested in the number of empty boxes—the problem is that discussed by both Romanovsky and Stevens, with the exception of an allowance for cases where the nuclear gene was not lost. We shall return to this problem also after discussing the general results.

5. The case of large numbers. In case the number of "balls" and "boxes" is large, it is natural and has been customary in similar problems to replace discrete variables by continuous, and derive differential equations. The process runs as follows: Let $y_0, y_1, y_2, \cdots, y_b, \cdots$ be the *fractions* of the total number of boxes containing no, one, two, $\cdots$, b, $\cdots$ balls. Let t be the average number of balls per box (artificially made continuous, so that we may, for example, have a total of $13 + 3\pi$ balls). Increase t to $t + dt$, then of the y_0 boxes previously containing no balls, $y_0\,dt$ will receive one. Of the y_1 boxes previously containing one ball each, $y_1\,dt$ will receive a second, and so on. Hence

$$\frac{dy_0}{dt} = -y_0,$$
$$\frac{dy_1}{dt} = y_0 - y_1,$$
$$\cdots$$

$$\frac{dy_b}{dt} = y_{b-1} - y_b,$$

$$\cdots$$

and if we start, when $t = 0$, with $y_0 = 1$, and $y_b = 0$ for $b > 0$, we find

$$y_b = \frac{t^b}{b!} e^{-t}, \qquad b = 0, 1, 2, \cdots. \tag{1}$$

The usefulness of this result has sometimes been in doubt, thus Opatowski [2, p. 164] says in a similar connection: "Consequently $\cdots$ the theory appears less accurate for small values of t"

It is shown in Part II that; where n_b boxes out of the total of N contain exactly b balls: (I) When the number of *balls and boxes* is large and fixed, (1) is a good approximation to the expectation of n_b/N. (II) When the total number of balls has a Poisson distribution, and t is interpreted as the expected number, (1) *reproduces the expectation* exactly. Since it is appropriate in most problems involving chemical reactions or irradiation to take the number of balls as having a

TABLE 1

A fixed or binomial number of balls and equal boxes

HYPOTHESIS

A total of m balls are independently distributed into N boxes or elsewhere, the chance of a particular ball entering a particular box is p. The number of boxes each containing exactly b balls is n_b.

$$\text{Mean of } n_b = E(n_b) = N\binom{m}{b}(1-p)^m\left(\frac{p}{1-p}\right)^b$$

$$\text{Variance of } n_b = E(n_b)(1 - (1 - \Phi(b, b))E(n_b))$$

$$\text{Covariance of } n_b \text{ and } n_c = -(1 - \Phi(b, c))E(n_b)E(n_c)$$

$$\Phi(b, c) = \left(1 - \frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}}\left(1 - \left(\frac{p}{1-p}\right)^2\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}$$

where $m^{(b)} = m(m-1)\cdots(m-b+1)$ involving b factors

Higher moments See Section 14

$$\text{Mean of } n_0 = N(1-p)^m$$

$$\text{Mean of } n_1 = Nm(1-p)^m\left(\frac{p}{1-p}\right)$$

$$\text{Variance of } n_0 = N(1-p)^m - N^2(1-p)^{2m} + N(N-1)(1-2p)^m$$

$$\text{Variance of } n_1 = N(N-1)m(m-1)(1-2p)^{m-2}p^2 + Nm(1-p)^{m-1}p - N^2m^2(1-p)^{2m-2}p^2$$

$$\text{Covariance of } n_0 \text{ and } n_1 = N(N-1)m(1-2p)^{m-1}p - n^2m(1-p)^{2m-1}p$$

Poisson distribution, the caution suggested by (I) is often shown unnecessary by (II). For this type of problem the differential equation is entirely adequate!

It is further shown in Part II that, in the Poisson case, the second moments are exactly those which correspond to random sampling from an infinite population with the fractions indicated by the mean number of boxes with 0, 1, 2, $\cdots$, b, $\cdots$ balls. This result is not accidental, and it is shown in Part III how we can see directly that the whole distribution in this case is that of random sampling from such a population.

6. The case of small numbers. The results of Part II also allow us to state the means, variances, and covariances, for the cases where the differential equations do not apply. The results are set forth in the following tables: Tables 1 and 2 apply to the cases where m balls are distributed among the given boxes and possibly others. Thus the total number of balls in the given boxes is either fixed, when there are no other boxes, or follows a binomial distribution.

TABLE 2

A fixed or binomial number of balls and unequal boxes

HYPOTHESIS

A total of m balls are independently distributed into N boxes or elsewhere, the chance of a particular ball entering the ith box being p_i. The average of the $p_i = p$. The sum of the squared fractional deviations of p_i from p is Λ. $p_i = p(1 + \lambda_i)$, $\Sigma_i \lambda_i^2 = \Lambda$. Terms in $\Sigma_i \lambda_i^3$, $\Sigma_i \lambda_i^4$, etc, are to be neglected. The number of boxes each containing exactly b balls is n_b.

Mean of $n_b = E(n_b) = N \binom{m}{b} (1 - p)^{m-b} p^b$ times

$$\left\{\left(1 + \frac{\Lambda}{2N(1-p)^2}\right)((mp - b)^2 - (m - b)p^2 - b(1 - p)^2)\right\}$$

Variances and covariances as in Table 1, using

$$\Phi(b, c) \approx \left(1 - \frac{1}{N}\right) \frac{(m - c)^{(b)}}{m^{(b)}} \left(1 - \left(\frac{p}{1-p}\right)^2\right)^m \left(\frac{1-p}{1-2p}\right)^{b+c} \left(1 + \frac{\Lambda\psi}{2N}\right)$$

where $\psi = 2bc\left(2p - \frac{1}{N}\right) +$ terms in p^2 and in $\frac{p}{N}$

The exact value of ψ is given in Section 16.

Mean of $n_0 = N(1 - p)^m \left(1 + \frac{\Lambda p^2 m(m-1)}{2N(1-p)^2}\right)$

Mean of $n_1 = Nm(1 - p)^{m-1} p \left(1 + \frac{\Lambda(m-1)p(1-mp)}{2N(1-p)^2}\right)$

TABLE 3

Poisson balls and equal boxes

HYPOTHESIS

A number of balls with the Poisson distribution, and expectation Nt are independently placed in N boxes. The number of boxes each containing exactly b balls is n_b.

Mean of $n_b = E(n_b) = N \frac{t^b}{b!} e^{-t}$

Variance of $n_b = N\left(\frac{t^b}{b!} e^{-t}\right)\left(1 - \frac{t^b}{b!} e^{-t}\right)$

Covariance of n_b and $n_c = -N\left(\frac{t^b}{b!} e^{-t}\right)\left(\frac{t^c}{c!} e^{-t}\right)$

Mean of $n_0 = Ne^{-t}$,
Mean of $n_1 = Nte^{-t}$,
Variance of $n_0 = Ne^{-t}(1 - e^{-t})$,
Variance of $n_1 = Nte^{-t}(1 - te^{-t})$,
Covariance of n_0 and $n_1 = -Nte^{-2t}$.

7. Discussion of the chemical problem. The number of oxygen molecules which have reacted in a given time is, at best, distributed Poisson Thus the differential equations would give the expected number of cuts, even if the number of balls or boxes were not large.

The fact that the numbers of balls and boxes, are large makes the variances and covariances so small as to be practically unimportant. Thus, for example, with $N = 10^{18}$, $t = 1$ (1 break per chain), we have:

$$\text{mean of } n_0 = \frac{1}{e} \times 10^{18},$$

$$\text{mean of } n_1 = \frac{1}{e} \times 10^{18},$$

$$\text{variance of } n_0 = \frac{1}{e}\left(1 - \frac{1}{e}\right) \times 10^{18},$$

$$\text{variance of } n_1 = \frac{1}{e}\left(1 - \frac{1}{e}\right) \times 10^{18},$$

$$\text{covariance of } n_0 \text{ and } n_1 = -\frac{1}{e^2} \times 10^{18}.$$

Thus the standard deviations are less than 1 part in 100 million of the mean.

TABLE 4

Poisson balls and varied boxes

HYPOTHESIS

A number of balls with the Poisson distribution are independently placed in N unequal boxes. The expected number placed in the ith box is t_i. The average of the t_i is t, $t_i = t(1 + \lambda_i)$ and $\Sigma_i \lambda_i^2 = \Lambda$. Terms in $\Sigma_i \lambda_i^3$, $\Sigma_i \lambda_i^4$, etc. are to be neglected. The number of boxes each containing exactly b balls is n_b.

$$\text{Mean of } n_b = E(n_b) = N \frac{t^b}{b!} e^{-t} \left(1 + \frac{\Lambda}{2N} ((b - t)^2 - b)\right)$$

$$\text{Variance of } n_b = E(n_b) - \frac{1}{N}\left(1 + \frac{\Lambda}{2N} (b - t)^2)(E(n_b))^2\right)$$

$$\text{Covariance of } n_b \text{ and } n_c = -\frac{1}{N}\left(1 + \frac{\Lambda}{2N} ((b - t)(c - t))E(n_b)E(n_c)\right)$$

$$\text{Mean of } n_0 = Ne^{-t}\left(1 + \frac{\Lambda t^2}{2N}\right)$$

$$\text{Mean of } n_1 = Nte^{-t}\left(1 + \frac{\Lambda(t^2 - 2t)}{2N}\right)$$

$$\text{Variance of } n_0 = Ne^{-t}\left(1 + \frac{\Lambda t^2}{2N}\right) - Ne^{-2t}\left(1 + \frac{3\Lambda t^2}{2N}\right)$$

$$\text{Variance of } n_1 = Nte^{-t}\left(1 + \frac{\lambda(t^2 - 2t)}{2N}\right) - Nt^2e^{-2t}\left(1 + \frac{\Lambda(3t^2 - 6t\,1)}{2N}\right)$$

$$\text{Covariance of } n_0 \text{ and } n_1 = -Nt^2e^{-2t}\left(1 + \frac{\Lambda(3t^2 - 3t)}{2N}\right)$$

8. Discussion of the bacteriological example. Although this example came from an irradiation experiment, we are not entitled to jump to the Poisson case. The balls are not actions of radiation, but rather previously existing enzyme particles. The purpose of the radiation is merely to make a failure to hand down a particle obvious.

For simplicity, let us begin by assuming that the irradiation has been strong enough to knock out all the nuclear genes and none of the enzyme particles. We face the following problem: "If the m enzyme particles are divided by chance among 8 descendants, what should be the distribution of mutants, that is, of boxes with no balls?"

As far as mean and variance, we can answer this question from Table 1, with $N = 8$ and $p = \frac{1}{8}$.

The results are

$$\text{mean number of mutants} = E(n_0) = 8(\tfrac{7}{8})^m,$$

$$\text{variance of same} = 8(\tfrac{7}{8})^m - 64(\tfrac{7}{8})^{2m} + 56(\tfrac{6}{8})^m.$$

For small values of m we get the values tabled below:

TABLE 5

Blanks out of 8

m	mean	variance	$\text{mean}\left(1 - \frac{\text{mean}}{8}\right)$
0	8	0.000	0.000
1	7	0.000	0.875
2	6 125	.109	1.436
3	5.359	.262	1.769
4	4.689	417	1.941
5	4.103	.556	1.998
6	3.590	.666	1.979
7	3.142	.747	1.908
8	2 749	.799	1.804
9	2 405	.825	1.682
10	2.105	829	1.551
15	1.079	.663	934
20	0.554	.426	.515

We notice that the variance is substantially less than the mean.

Now it might be that the number of enzyme particles is not constant from bacterium to bacterium. It would not be unreasonable if it had a Poisson distribution. If this were the case, we would revert to the differential equation solution, which is also given in Table 3. The last column in Table 5 shows the variance which would then arise for the same means. The variance is still somewhat less than the mean. The situation is shown graphically in Figure 1.

If the actual distribution of n_0 is desired, then it can be calculated for the case where m is fixed from the tables in Stevens' paper [4], and when m is distributed Poisson it is merely a binomial distribution.

PART II

DERIVATIONS

9. The chance generating function. We are considering the following class of problems: "balls" are placed *independently* in "boxes" and then the number n_0 of empty compartments, the number n_1 of compartments containing exactly

one ball, $\cdots$, the number n_b of boxes with exactly b balls, and so on, are observed. We are interested in the moments of n_0, n_1, n_2, $\cdots$, n_b, $\cdots$ both simple and mixed.

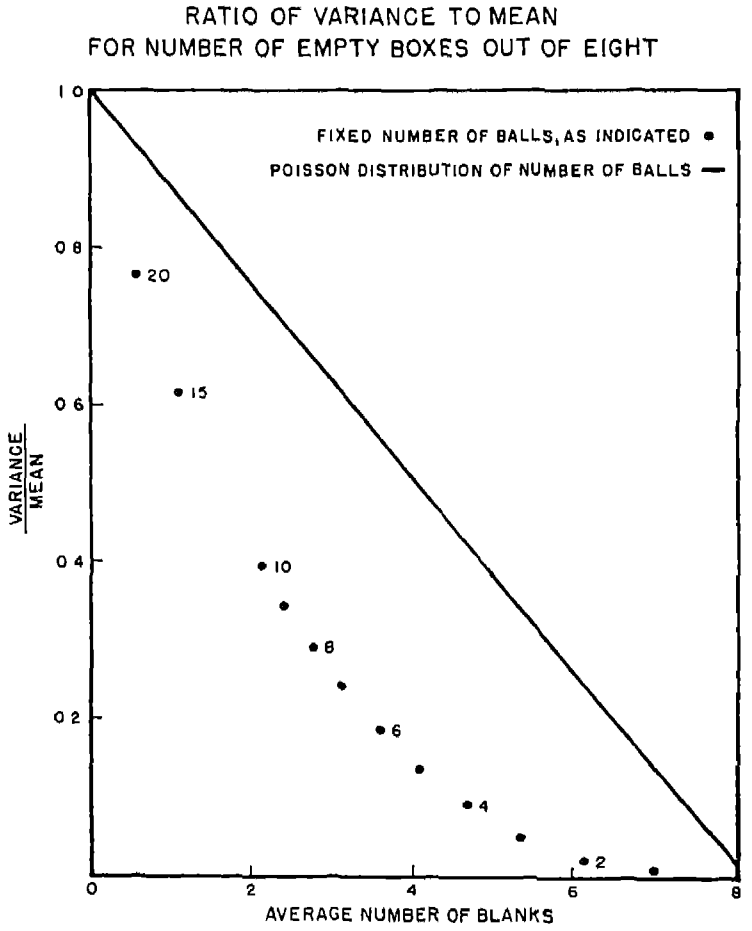

Figure 1

We define chance quantities x_{iq} by

$$x_{iq} = \begin{cases} x, & q\text{th ball in the } i\text{th box,} \\ 1, & \text{otherwise.} \end{cases}$$

Clearly the product of all x_{iq} for fixed i is given by

$$\Pi_q x_{iq} = x^{(\text{number of balls in the } i\text{th box})}$$

Thus $\Pi_q x_{iq} = x^b$ if and only if there are exactly b balls in the ith box. Hence the coefficient of x^b in $\Sigma_i \Pi_q x_{iq}$, the sum of $\Pi_q x_{iq}$ over all boxes i, is n_b, the number of boxes containing exactly b balls.

We have the relation

$$\Sigma_b n_b x^b = f(x) = \Sigma_i \Pi_q x_{iq},$$

where $f(x)$ is a chance function, and the n_b and the x_{iq} are chance quantities.

Now we take expectations of both sides, and use the fact that the expectation of a sum is the sum of the expectations to obtain

$$\Sigma_b x^b E(n_b) = E(f(x)) = \Sigma_i E(\Pi_q x_{iq})$$

Now x_{iq} and x_{ir}, for $q \neq r$, are independent since they are determined by different and independent balls. Hence $E(\Pi_q x_{iq}) = \Pi_q(E x_{iq})$ and we have the basic formula

$$E(f(x)) = \Sigma_b x^b E(n_b) = \Sigma_i \Pi_q E(x_{iq}). \tag{1}$$

10. Higher moments. By extending this device, we can obtain generating functions for higher moments. Instead of the x_{iq}, we introduce a whole sequence of chance quantities $x_{iq}, y_{iq}, z_{iq}, \cdots, w_{iq}$, defined by

$$(x_{iq}, y_{iq}, \cdots, w_{iq}) = \begin{cases} (x, y, \cdots, w), & q\text{th ball in } i\text{th box}, \\ (1, 1, \cdots, 1), & \text{otherwise}. \end{cases}$$

We find immediately that

$$\begin{aligned} f(x)f(y)\cdots f(w) &= (\Sigma_i \Pi_q x_{iq})(\Sigma_j \Pi_r y_{jr}) \cdots (\Sigma_n \Pi_p w_{np}) \\ &= \Sigma_i \Sigma_j \cdots \Sigma_n \Pi_q x_{iq} y_{jq} \cdots w_{nq}. \end{aligned}$$

Taking expectations on both sides

$$\begin{aligned} E(f(x)f(y)\cdots f(w)) &= \Sigma_i \Sigma_j \cdots \Sigma_n E(\Pi_q x_{iq} y_{jq} \cdots w_{nq}) \\ &= \Sigma_i \Sigma_j \cdots \Sigma_n \Pi_q E(x_{iq} y_{jq} \cdots w_{nq}), \end{aligned}$$

where we have used the fact that $x_{iq} y_{jq} \cdots w_{nq}$ and $x_{ir} y_{jr} \cdots w_{nr}$ are independent when $q \neq r$ since they are determined by different and independent balls

On the other hand,

$$\begin{aligned} f(x)f(y)\cdots f(w) &= (\Sigma_b n_b x^b)(\Sigma_c n_c y^c) \cdots (\Sigma_a n_a w^a) \\ &= \Sigma_b \Sigma_c \cdots \Sigma_a (n_b n_c \cdots n_a)(x^b y^c \cdots w^a) \end{aligned}$$

so that

$$E(f(x)f(y)\cdots f(w)) = \Sigma_b \Sigma_c \cdots \Sigma_a (x^b y^c \cdots w^a) E(n_b n_c \cdots n_a).$$

Equating the two expressions for the expectation of $f(x)f(y)\cdots f(w)$, we have, finally, the generating function for $E(n_b n_c \cdots n_a)$ in the form

$$\sum_{b,c,\cdots,a} (x^b y^c \cdots w^a) E(n_b n_c \cdots n_a) = \sum_{i,j,\cdots,n} \Pi_q E(x_{iq} y_{jq} \cdots w_{nq}). \tag{2}$$

Thus a knowledge of $E(x_{iq} y_{jq} \cdots w_{nq})$ will allow us to determine the moments of the n's.

11. A fixed or binomial number of balls and equal boxes. Let there be N boxes, and m balls, each with probability p of entering each box. If $pN = 1$ we have the case where m balls always appear in the boxes taken together—the case of a fixed number of balls. If $pN < 1$, the number of balls appearing in all boxes taken together is a binomial with expectation mpN.

Now x_{iq} equals 1 with probability $1 - p$ and equals x with probability p, hence (1) becomes

$$\Sigma_b x^b E(n_b) = \Sigma_i \Pi_q (1 - p + px) = N(1 - p + px)^m.$$

Using the binomial theorem, the coefficient of x^b is

$$E(n_b) = N\binom{m}{b}(1-p)^{m-b}p^b = N\binom{m}{b}(1-p)^m\left(\frac{p}{1-p}\right)^b. \tag{3}$$

Now if p is small, we may approximate $1 - p$ by e^{-p} and by 1, respectively, in its two occurences, where

$$E(n_b) \approx N\binom{m}{b}e^{-mp}p^b$$

and if m is large compared to b this becomes

$$E(n_b) \approx N\frac{(mp)^b}{b!}e^{-mp}.$$

12. Second moments. We must study $E(x_{iq}y_{jq})$. If $i = j$ then this is $(1 - p + pxy)$ since the qth ball falls into *both* the ith and jth boxes with probability p, otherwise into neither. If $i \neq j$, we immediately find the expectation to be $(1 - 2p + px + py)$.

Hence, since $i = j$ in N cases, and $i \neq j$ in $N(N - 1)$ cases,

$$\Sigma_{ij}\Pi_q E(x_{iq}y_{jq}) = N(1 - p + pxy)^m + N(N-1)(1 - 2p + px + py)^m,$$

by (2) this equals $\Sigma_{bc}x^b y^c E(n_b n_c)$, and using the multinomial expansion we find

$$E(n_b n_c) = N(N-1)\binom{m}{b\,c}(1-2p)^{m-b-c}p^{bc} + \delta(b,c)N\binom{m}{b}(1-p)^{m-b}p^b,$$

where $\delta(b, c) = 1$ when $b = c$ and is zero otherwise, and where the multinomial coefficient $\binom{m}{b\,c}$ is given by

$$\binom{m}{b\,c} = \frac{m!}{b!c!(m-b-c)!}.$$

We now set

$$E(n_b n_c) = E(n_b)E(n_c)\Phi(b,c) + \delta(b,c)E(n_b), \tag{4}$$

when

$$\Phi(b, c) = \frac{N(N-1)\binom{m}{b\,c}(1-2p)^m\left(\frac{p}{1-2p}\right)^{b+c}}{N\binom{m}{b}(1-p)^m\left(\frac{p}{1-p}\right)^b N\binom{m}{c}(1-p)^m\left(\frac{p}{1-p}\right)^c}$$

$$= \left(1-\frac{1}{N}\right)\frac{\binom{m}{b\,c}}{\binom{m}{b}\binom{m}{c}}\left(\frac{1-2p}{(1-p)(1-p)}\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c} \tag{5}$$

$$= \left(1-\frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}}\left(1-\left(\frac{p}{1-p}\right)^2\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}$$

where $u^{(b)} = u(u-1)\cdots(u-b+1)$ denotes a descending factorial with b factors.

Notice that, if the n_b were independently distributed in Poisson distributions, the second moments would be given by the same formula with $\Phi(b, c) = 1$, while if they were distributed like a multinomial sample from an infinite population the second moments would be given by the same formula with $\Phi(b, c) = 1 - \frac{1}{N}$.

For small p, we have

$$\Phi(b, c) \approx \left(1-\frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}},$$

and if m is large compared to b and c, this approaches the multinomial value

$$\Phi(b, c) \approx \left(1-\frac{1}{N}\right).$$

13. Variances and covariances. The variances and covariances are given by

$$\begin{aligned}\text{Variance } (n_b) &= E(n_b n_b) - E(n_b)E(n_b)\\ &= E(n_b)(1-(1-\Phi(b, b))E(n_b)),\end{aligned}$$

and

$$\text{Covariance } (n_b, n_c) = -(1-\Phi(b, c))E(n_b)E(n_c).$$

Thus the covariance of n_b and n_c will vanish when, and only when $\Phi(b, c) = 1$.

Let us suppose $pN = \frac{1}{\beta}$, with p small and m and N large, and see if $\Phi(b, c)$ can be unity. Since a preliminary calculation shows it to be reasonable, let us put $m = \gamma N$. Then

$$\Phi(b, c) \approx (1-\beta p)\frac{(\gamma N - c)^{(b)}}{(\gamma N)^{(b)}}(1-p^2)^{\gamma N}(1+p)^{b+c}.$$

An easy calculation shows that the ratio of descending factorials is nearly

$$e^{-bc/\gamma N} = e^{(-bc\beta/\gamma)p},$$

making further natural approximations,

$$\ln \Phi(b, c) \approx -\beta p - \frac{bc\beta}{\gamma} p - \gamma N p^2 + (b + c)p$$

and this may be written

$$\ln \Phi(b, c) \approx -\frac{\beta p}{4\gamma}\left(\left(2\frac{\gamma}{\beta} - b - c + \beta\right)^2 + 4\beta c - (b - \beta - c)^2\right),$$

and this vanishes for real γ when and only when $| b - \beta - c | \geq \sqrt{4\beta c}$. This, then, is the condition on b and c which permits the existence of two ratios of m to N so that for either ratio and large N there will be no correlation between n_b and n_c.

14. Higher moments. To deal with the third moments, we need $E(x_{iq}y_{jq}z_{kq})$, which is easily seen to behave as follows:

Relation of ijk	*number of occurrences*	*Expectation of* $x_{iq}y_{jq}z_{kq}$
$i = j = k$	N	$1 - p + pxyz$
$i = j \neq k$	$N(N - 1)$	$1 - 2p + pxy + pz$
$i = k \neq j$	$N(N - 1)$	$1 - 2p + pxz + py$
$j = k \neq i$	$N(N - 1)$	$1 - 2p + pyz + px$
different	$N(N - 1)(N - 2)$	$1 - 3p + px + py + pz$

Thus we have

$$\Sigma_{bcd}x^b y^c z^d E(n_b n_c n_d) = N(1 - p + pxyz)^m + N(N - 1)(1 - 2p + pxy + pz)^m$$
$$+ N(N - 1)(1 - 2p + pxz + py)^m + N(N - 1)(1 - 2p + pyz + px)^m$$
$$+ N(N - 1)(N - 2)(1 - 3p + px + py + pz)^m$$

from which we can calculate all third moments.

In general if ϵ is a decomposition of the product $xyz \cdots w$ into α monomials $u_1, u_2, \cdots, u_\alpha$, where order is disregarded (for example: $xyz = (xz)y = (zx)y = y(zx) = y(xz)$ is a single decomposition with $\alpha = 2$, $u_1 = xz$, $u_2 = y$), then the generating function becomes

$$\Sigma_\epsilon N^{(\alpha)}(1 + (u_1 + u_2 + \cdots + u_\alpha - \alpha)p)^m.$$

15. Poisson balls and equal boxes. To reach a Poisson distribution we let $m \to \infty$ and $p \to 0$ so that $mNp = tN$, where t is the *average number of balls per box* in the Poisson distribution.

Since

$$p^b\binom{m}{b} \to \frac{t^b}{b!}$$

under these conditions, (3) becomes

$$E(n_b) = N\frac{t^b}{b!}e^{-t} \tag{6}$$

and from (5) it follows that the limit of $\Phi(b, c)$ is $\left(1 - \frac{1}{N}\right)$ so that

$$E(n_b n_c) = N(N-1)\frac{t^{b+c}}{b!\,c!}e^{-2t} + \delta(b, c)N\frac{t^b}{b!}e^{-t}, \tag{7}$$

and hence

$$\text{Variance } (n_b) = N\left(\frac{t^b}{b!}e^{-t}\right)\left(1 - \frac{t^b}{b!}e^{-t}\right), \tag{8}$$

$$\text{Covariance } (n_b, n_c) = -N\left(\frac{t^b}{b!}e^{-t}\right)\left(\frac{t^c}{c!}e^{-t}\right). \tag{9}$$

Notice that these are the moments of the numbers of objects of types b, c, $\cdot$, in a random sample of N from an infinite population where the fraction of b's is $t^b e^{-t}/b!$, just as it should be.

16. Fixed or binomial balls and varied boxes. We now consider the case where the chance of any ball entering the ith box is p_i. We shall again not restrict ourselves to the case $\Sigma_i p_i = 1$

The expectation of x_{iq} is immediately seen to be $(1 + p_i(x - 1)) = (1 - p_i + p_i x)$, so that the generating function is

$$f(x) = \Sigma_i(1 - p_i + p_i x)^m$$

and the expectation of n_b is

$$E(n_b) = \binom{m}{b}\Sigma_i(1 - p_i)^{m-b}p_i^b = \binom{m}{b}\Sigma_i(1 - p_i)^m\left(\frac{p_i}{1 - p_i}\right)^b. \tag{10}$$

Following Stevens [4] with a slight modification, let us set $p_i = p(1 + \lambda_i)$, where p is the average of the p_i, so that $\Sigma_i\lambda_i = 0$. Then

$$(1 - p_i) = (1 - p(1 + \lambda_i)) = (1 - p)\left(1 - \frac{p\lambda_i}{1 - p}\right),$$

so that

$$\Sigma_i(1 - p_i)^{m-b}p_i^b = (1 - p)^{m-b}p^b\,\Sigma_i\left(1 - \frac{p\lambda_i}{1 - p}\right)^{m-b}(1 + \lambda_i)^b.$$

Expanding the summand, we find

$$1 + \left\{ -\frac{(n-b)p}{1-p} + b \right\} \lambda_i + \left\{ \frac{(m-b)(m-b-1)p^2}{2(1-p)^2} - \frac{(m-b)bp}{1-p} + \frac{b(b-1)}{2} \right\} \lambda_i^2 + O(\lambda_i^3).$$

Hence, setting $\Sigma_i \lambda_i^2 = \Lambda$ (notice this is not the same as Stevens' Λ!), we have

$$E(n_b) = \binom{m}{b} (1-p)^{m-b} p^b \left\{ N + \tfrac{1}{2}\Lambda \left(\frac{m-b}{m-b-1} \frac{(p(m-1)-b)^2}{(1-p)^2} - \frac{b(m-1)}{m-b-1} \right) \right\} + O(\Sigma_i \lambda_i^3).$$

The expectation for all $p_i = p$ has been modified by multiplication by

$$1 + \frac{\Lambda}{2N} \left\{ \frac{m-b}{m-b-1} \frac{(p(m-1)-b)^2}{(1-p)^2} - \frac{b(m-1)}{m-b-1} \right\} \tag{11}$$

plus terms of higher order. For large N and consequently small p the quantity in braces is nearly

$$b\left(b - \frac{m}{m-b}\right)$$

and more roughly is approximately b^2. Similarly, the expectations of second moments are

$$E(n_b n_c) = \binom{m}{b\,c} \sum_{i \neq j} (1 - p_i - p_j)^{m-b-c} p_i^b p_j^c + \delta(b,c) \binom{m}{b} \sum_i (1-p_i)^{m-b} p_i^b,$$

whence

$$\Phi(b,c) = \frac{\binom{m}{b\,c} \sum_{i \neq j} (1 - p_i - p_j)^{m-b-c} p_i^b p_j^c}{\binom{m}{b}\binom{m}{c} \sum_i (1-p_i)^{m-b} p_i^b \sum_j (1-p_j)^{m-c} p_j^c}. \tag{12}$$

Making the same sort of expansion yields

$$\Phi(b,c) \approx \left(1 - \frac{1}{N}\right) \frac{(m-c)^{(b)}}{m^{(b)}} \left(1 - \frac{p^2}{(1-p)^2}\right)^m \left(\frac{1-p}{1-2p}\right)^{b+c} \left(1 + \frac{\Lambda\psi}{2N}\right) \tag{13}$$

where terms in $\Sigma\lambda_i^3$ have been neglected (note that

$$\sum_{i \neq j} \lambda_i \lambda_j = -\sum_i \lambda_i^2 = -\Lambda.),$$

and where

$$\psi = \left\{ \frac{m-b-c}{m-b-c-1} \frac{N-2}{N-1} (1-2p)^{-2} - \frac{m-b}{m-b-1} (1-p)^{-2} \right\} \cdot \{p(m-1)-b\}^2$$

$$+\left\{\frac{m-b-c}{m-b-c-1}\,\frac{N-2}{N-1}\,(1-2p)^{-2}-\frac{m-c}{m-c-1}\,(1-p)^{-2}\right\}$$
$$\cdot\{p(m-1)-c\}^2$$
$$+\frac{1}{2}\,\frac{m-b-c}{m-b-c-1}\left\{\frac{N}{N-1}-\frac{N-2}{N-1}\,(1-2p)^{-2}\right\}(b-c)^2$$
$$+\frac{1}{m-b-c-1}\left\{\frac{2bc}{N-1}-\frac{b^2c}{m-b-1}-\frac{c^2b}{m-c-1}\right\}.$$

This can be reduced to

$$\psi = 2bc\left(2p-\frac{1}{N}\right)+O(p^2)+O\left(\frac{b}{N}\right),$$

and for $p = 1/N + O(p^2) + O\left(\frac{b}{N}\right)$.

$$\psi = 2pbc + O(p^2).$$

17. Poisson balls and varied boxes. To reach the Poisson limit, we let $m \to \infty$ and $p_i \to 0$ so that $mp_i = t_i$. The generating function for first moments becomes

$$f(x) = \Sigma_i e^{-t_i+t_ix}$$

and the expectation of n_b is

$$E(n_b) = \Sigma_i \frac{t_i^b}{b!}\,e^{-t_i}. \tag{15}$$

If we set $t_i = t(1+\Lambda_i)$, this becomes

$$E(n_b) = \frac{t^b}{b!}\,e^{-t}\,\Sigma_i(1+\lambda_i)^b e^{-t\lambda_i}$$

The summand expands in the form

$$\left(1+b\lambda_i+\frac{b(b-1)}{2}\lambda_i^2+\frac{b(b-1)(b-2)}{6}\lambda_i^3+\cdots\right)$$
$$\times\left(1-t\lambda_i+\frac{t^2}{2}\lambda_i^2-\frac{t^3}{6}\lambda_i^3+\cdots\right)$$
$$= 1+(b-t)\lambda_i+\left(\frac{b(b-1)}{2}-bt+\frac{t^2}{2}\right)\lambda_i^2+\cdots.$$

If t is chosen as the average of the t_i so that $\Sigma\lambda_i = 0$, the sum becomes

$$N+\left(\frac{(b-t)^2-b}{2}\right)\Sigma\lambda_i^2+\left(\frac{(b-t)^3}{6}-\frac{3b-2}{6}+\frac{bt}{2}\right)\Sigma\lambda_i^3+\cdots.$$

Again setting $\Sigma\lambda_i^2 = \Lambda$ we have

$$E(n_b) \approx \frac{t^b}{b!}\,e^{-t}\left(N+\left(\frac{(b-t)^2-b}{2}\right)\Lambda\right) \tag{16}$$

which can be written

$$E(n_b) \approx N \frac{t^b}{b!} e^{-t} \left(1 + \frac{\Lambda}{2N} ((b - t)^2 - b)\right).$$

The generating function for the second moments is

$$f(x)f(y) = \Sigma_{ij} e^{-t_i + t_i x} \, {}^{t_j + t_j y}$$

so that the expectation of $n_b n_c$ is

$$E(n_b n_c) = \sum_{i \neq j} \frac{t_i^b t_j^c e^{-t_i - t_j}}{b!c!} + \delta(b, c) \sum_i \frac{t_i^b}{b!} e^{-t_i} \tag{17}$$

which becomes

$$E(n_b n_c) = \frac{t^{bc}}{b!c!} e^{-2t} \sum_{i \neq j} (1 + \lambda_i)^b (1 + \lambda_j)^c e^{-t\lambda_i - t\lambda_j} + \delta(b, c) E(n_b),$$

whence we can derive

$$\Phi(b, c) \approx 1 - \frac{1}{N} - \frac{\Lambda}{2N^2} (b - t)(c - t). \tag{18}$$

Thus

$$\text{Variance } (n_b) \approx E(n_b) - \frac{1}{N} \left\{1 + \frac{\Lambda}{2N} (b - t)^2\right\} (E(n_b))^2, \tag{19}$$

$$\text{Covariance } (n_b n_c) \approx -\frac{1}{N} \left(1 + \frac{\Lambda}{2N} (b - t)(c - t)\right) E(n_b) E(n_c). \tag{20}$$

18. Boxes in a systematic square. Another case which it may be worthwhile to write down arises when the boxes are systematically "rotated" under "spouts" of different probability. That is, the number of balls m is a multiple of the number of boxes N, and the probability of the qth ball entering the ith box depends on the value of $q - i$ taken modulo N. An example for $N = 3$ and $m = 6$ follows:

Probabilities of entry

Box	Ball 1	2	3	4	5	6
1	p_0	p_1	p_2	p_0	p_1	p_2
2	p_2	p_0	p_1	p_2	p_0	p_1
3	p_1	p_2	p_0	p_1	p_2	p_0

If $m = kN$ and the subscript r runs through 0, 1, 2, $\cdots$, $N - 1$, then the expectation of $f(x)$ becomes

$$\Sigma_i \Sigma_q E(x_{iq}) = N\{\Pi_r(1 - p_r + p_r x)\}^m.$$

Thus first moments, and by proceeding similarly higher moments, are available for this case also.

PART III

THE POISSON CASE

19. The Poisson case with equal boxes. The Poisson case is obtained in the limit as $m \to \infty$ and $p \to 0$ with $pm = t$ We wish to show that, in the limit, the number of balls in the different boxes are independent. Let $k_1, k_2, \cdots, k_N$ be the number of balls in the first, second, $\cdots$, Nth box, respectively. Then the distribution of the k's is given by, where we write $k = k_1 + k_2 + \cdots + k_N$,

$$\frac{m^{(k)}}{k_1!k_2! \cdots k_N!} p^k(1 - Np)^{m-k} = \frac{m^{(k)}}{m^k} \frac{(1 - Np)^{m-k}}{e^{-Nmp}} \Pi_i \frac{(mp)^{k_i} e^{-mp}}{k_i!}$$

Now the first two fractions clearly approach unity in the limit, and the independence is proved.

Since the number of balls in each box has an independent Poisson distribution, the distribution of the numbers of boxes each with exactly b balls is that of a random sample of N from an infinite population—namely it is a multivariate distribution with probabilities

$$\frac{(mp)^b e^{-mp}}{b!} .$$

REFERENCES

[1] H. H. Newcombe, "Delayed phenotypic expression of spontaneous mutations in *Escherichia coli*," *Genetics*, *Vol 33* (1948), p 447-476

[2] I. Opatowski, "Chain processes and their biophysical applications: Part I. General Theory," *Bulletin of Mathematical Biophysics*, *Vol 7* (1945), p. 161-180

[3] V. Romanovsky, "Su due problemi di distribuzione casuale," *Giornale dell'Istituto Italiano degli Attuari*, *Vol 5* (1934), p. 196-218.

[4] W. L. Stevens, "Significance of grouping," *Annals of Eugenics*, *Vol 8* (1937-1938), p. 57-59.

THE POWER OF THE CLASSICAL TESTS ASSOCIATED WITH THE NORMAL DISTRIBUTION

By J. Wolfowitz

Columbia University

Summary. The present paper is concerned with the power function of the classical tests associated with the normal distribution. Proofs of Hsu, Simaika, and Wald are simplified in a general manner applicable to other tests involving the normal distribution. The set theoretic structure of several tests is characterized. A simple proof of the stringency of the classical test of a linear hypothesis is given.

1. Introduction. The present paper is concerned with the optimum properties, from the power function viewpoint, of the classical tests associated with the normal distribution. In 1941 Hsu [2] proved the result stated in Section 2 below, which is concerned with the general linear hypothesis (in this connection his paper [1] of 1938 will be of interest). Also in 1941 Simaika [3] proved similar results for the tests based on the multiple correlation coefficient and Hotelling's generalization of Student's t. In 1942, Wald [4] gave a generalization of Hsu's result.

In the present paper we give short and simple proofs of almost all these results, and a simple proof of the stringency property of the analysis of variance (Section 5). These proofs rest on theorems which characterize the set theoretic structure of the tests. Thus, while the proofs of Hsu, Simaika and Wald are rather elaborate and each problem is essentially attacked *de novo*, the methods of the present paper are in effect applicable to the classical tests based on the normal distribution. For these tests it will not be difficult to demonstrate the analogues of Theorems 1 and 3, and of the results of Hsu, Simaika, and Wald. In the present paper we first treat the general linear hypothesis, because it is the simplest problem, its solution is easiest to describe, and it admits Wald's integration theorem. Multivariate analogues of the latter are rather artificial and not as simple. We then discuss the problem of the multiple correlation coefficient, because it seems to be more difficult than that of Hotelling's T and indeed, to include all the essential multivariate difficulties. Theorems 6 and 7 are the analogues of 1 and 3, respectively, while Theorem 9 describes the essential property of the power function which is of interest to us. In other multivariate problems one will prove the analogues of Theorems 6, 7 and 9. A generally inclusive formulation is no doubt possible. Theorems 5 and 9 are slightly more general than the theorems of Hsu and Simaika.

Many of the statements below may be not valid on exceptional sets of measure zero. Usually this is so stated, but sometimes, for reasons of brevity or to avoid repetition, this qualification may be omitted. The reader will have no difficulty supplying it wherever necessary.

The author is indebted to Erich L. Lehmann of the University of California, who carefully read a first version of this paper. Theorem 4 below was arrived at independently by Professor Lehmann, with a somewhat different proof

2. The general linear hypothesis. In canonical form the general linear hypothesis may be stated as follows: The chance variables

$$X_1, X_2, \cdots, X_{k+l}$$

have at $x_1, \cdots, x_{k+l}$, the density function

$$(2.1) \qquad (\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2 + \sum_{k+1}^{k+l} x_i^2\right\}\right] = f(\eta, \sigma)$$

with $\sigma, \eta_1, \cdots, \eta_k$ all unknown.

Let η be the vector $(\eta_1, \cdots, \eta_k)$. The null hypothesis H_0 states that

$$\eta_1 = \cdots = \eta_k = 0$$

and is to be tested with constant size $\alpha < 1$ (identically in σ).

Let D be any admissible critical region for testing H_0. If A is any event let $P\{A \mid \eta, \sigma\}$ denote the probability of A when η and σ are the parameters of (2.1) We have then

$$P\{D \mid 0, \sigma\} = \alpha$$

identically in σ, where 0 is the vector with k components all of which are zero. We now prove a property which characterizes all D. This theorem is due to Neyman and Pearson [12], and is given here only for completeness.

THEOREM 1. *The fraction of the surface area of the sphere*

$$\sum_1^{k+l} x_i^2 = c^2$$

which lies in D is α for almost all c

PROOF. Let a be any positive integer, h a positive parameter, and $\psi(y)$ a measurable function of y defined for $y > 0$ and such that $0 \leq \psi(y) \leq 1$. In view of the distribution of ΣX_i^2, it will be enough to prove that, if

$$\frac{h^{a+1}}{\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = \alpha$$

identically for all positive h, that then

$$\psi(y) = \alpha \text{ for almost all } y.$$

Write

$$(2.2) \qquad \frac{1}{\alpha\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = h^{-(a+1)}.$$

Differentiating both members k times with respect to h and then setting $h = 1$

we obtain the following result. The function

$$\frac{1}{\alpha\Gamma(a+1)}\,\psi(y)y^a e^{-y}$$

is a density function with kth moment

$$\mu_k = (a+1)(a+2)\cdots(a+l).$$

The moments μ_k are the moments of the density function

$$\frac{1}{\Gamma(a+1)}\,y^a e^{-y}.$$

They satisfy the Carleman criterion [5, p. 19, Th.1.10], and hence no essentially different distribution can have these moments. This proves the desired result.

THEOREM 2 (Wald). *Among all tests of the general linear hypothesis the analysis of variance test has the property that, for all positive d, the integral of its power on the surface $\eta^2 = d^2$ is a maximum.*

PROOF. Let c be any positive number. We have only to show that if we allocate to the critical region D of the test the fraction α of the surface area of the sphere

$$\sum_1^{k+l} x_i^2 = c^2 \tag{2.3}$$

for which

$$C = \frac{\sum_1^k x_i^2}{\sum_{k+1}^{k+l} x_i^2}$$

is as large as possible and that if we do this for all c, the desired maximum of the integral of the power will be achieved. If C is as large as possible so is

$$\frac{\sum_1^k x_i^2}{\sum_1^{k+l} x_i^2} = \frac{\sum_1^k x_i^2}{c^2}.$$

Let $a_1, \cdots, a_{k+l}$ be any point on the sphere (2.3). Let db be the differential of area on the surface $\eta^2 = d^2$. Then

$$\int\cdots\int_{\eta^2=d^2} f(\eta,\sigma)\,db = (\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left\{-\frac{(c^2+d^2)}{2\sigma^2}\right\} \cdot \int\cdots\int_{\eta^2=d^2} \exp\left\{\frac{(\eta)'z}{\sigma^2}\right\} db, \tag{2.4}$$

where z is the vector $(a_1, \cdots, a_k)$ and $(\eta)'z$ is the scalar product of the two vectors. This last integral is easily seen to depend only upon $|z|$ and to be monotonically increasing in $|z|$. This proves the theorem.

COROLLARY (Hsu). *Among all tests of the general linear hypothesis whose power is a function of η^2 only, the analysis of variance is the most powerful.*

3. The set theoretic structure of tests whose power is a function only of η^2/σ^2. Wald's result (Theorem 2) cannot always be extended, in its simple form, to tests involving the multivariate normal distribution, but this can be done with Hsu's theorem (corollary to Theorem 2) In order to see what is involved we shall investigate the set theoretic structure of tests of the general linear hypothesis whose power is a function only of η^2/σ^2.

Let $q(x_1, \cdots, x_k)$ be the set of points in the region D whose first k coordinates are $x_1, \cdots, x_k$. Let $A(x_1, \cdots, x_k, \sigma)$ be the integral of

$$(2\pi\sigma^2)^{-(l/2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{l} x_{k+j}^2\right\}\right]$$

with respect to $x_{k+1}, \cdots, x_{k+l}$, taken over $q(x_1, \cdots x_k)$. We first prove the following:

LEMMA. *Suppose the power of D is a function only of η^2/σ^2. Then for two points*

$$x_1, \cdots, x_k$$

and

$$x_1', \cdots, x_k'$$

such that

$$\sum_1^k x_i^2 = \sum_1^k x_i'^2 \tag{3.1}$$

we have

$$A(x_1, \cdots, x_k, \sigma) = A(x_1', \cdots, x_k', \sigma) \tag{3.2}$$

identically in σ, with the exception of a set of measure zero.

PROOF. Suppose the statement is false. Then under some orthogonal transformation T of $x_1, \cdots, x_k$ the region D would go over into a region D^* with the following property: Let $A^*(x_1, \cdots, x_k, \sigma)$ have the same definition for the region D^* as $A(x_1, \cdots, x_k, \sigma)$ has for D Then on a set of positive measure[1] we would have

$$A(x_1, \cdots, x_k, \sigma) \gtrless A^*(x_1, \cdots, x_k, \sigma). \tag{3.3}$$

We shall now show that (3.3) results in a contradiction. We have

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid T\eta, \sigma\} \tag{3.4}$$

identically in η. By the property of the region D, therefore, we have

$$P\{D \mid \eta, \sigma\} = P\{D \mid T^{-1}\eta, \sigma\}$$

[1] The situation here is similar to that described in footnote 3

and hence

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid \eta, \sigma\} \tag{3.5}$$

identically in η. Thus we obtain

$$\int (2\pi\sigma^2)^{-(k/2)} A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k \tag{3.6}$$

$$\equiv \int (2\pi\sigma^2)^{-(k/2)} A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k$$

with the integrations taking place over the entire space. Differentiating both members with respect to the components of η and setting $\eta = 0$, we obtain that the two density functions (for fixed σ)

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

and

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

have identical moments. We shall now argue that these moments satisfy the conditions of Cramér and Wold [7, Th. 2], so that the two density functions are essentially the same, in contradiction to (3.3). The Cramér-Wold theorem states the following: *Let $Y_1, \cdots, Y_k$ be k chance variables with a joint distribution function, and write*

$$\lambda_{2n} = \sum_{i=1}^k EY_i^{2n}.$$

Then the divergence of the series

$$\sum_{n=1}^{\infty} \lambda_{2n}^{-(1/2n)}$$

is sufficient to ensure that there exists essentially only one distribution which has these moments. We notice that the factor $1/\alpha$ of course makes no difference. If we set $A(x_1, \cdots, x_k, \sigma)$ and $A^*(x_1, \cdots, x_k, \sigma)$ both identically unity and consider the resulting moments which enter into the λ_{2n}, we see that these moments satisfy the Cramér-Wold condition. Now A and A^* are ≤ 1. Thus, using the true value of A can serve only to increase the value of $\lambda_{2n}^{-(1/2n)}$, so that the series will diverge a fortiori. This proves the lemma.

The following theorem helps to describe the set theoretic structure of tests whose power is a function only of $\lambda = \eta^2/\sigma^2$:

THEOREM 3. *Let D be a test whose power is a function only of λ. Let u be any positive number, and $D(x_1, \cdot\ \cdot, x_k, u)$ be the fraction of the "area" of the sphere $\Sigma_{i=1}^k x_{k+i}^2 = u^2$ occupied by points which are in D and whose first k coordinates are $x_1, \cdots, x_k$. If*

$$\sum_1^k x_i^2 = \sum_1^k x_i'^2 \tag{3.7}$$

then, except on a set of measure zero,

$$D(x_1, \cdots, x_k, u) = D(x_1', \cdots, x_k', u). \tag{3.8}$$

PROOF. We shall show that, if the power of D is a function only of λ, the failure of (3.7) to imply (3.8) would contradict the preceding lemma. Suppose then that (3.8) is not true on a set of positive measure. Under some orthogonal transformation on $x_1, \cdots, x_k$ we obtain[2] a function $D^*(x_1, \cdots, x_k, u)$ which differs from $D(x_1, \cdots, x_k, u)$ on a set of positive measure and such that, for almost every $x_1, \cdots, x_k$,

$$A(x_1, \cdots, x_k, \sigma) = K \int_0^\infty D(x_1, \cdots, x_k, u) \sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

$$= K \int_0^\infty D^*(x_1, \cdots, x_k, u) \sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

identically in σ, where K is a suitable constant of no interest to us. Multiplying by σ^l, differentiating repeatedly under the integral sign with respect to σ, and setting $\sigma = 1$, we obtain the result that the two density functions in u,

$$\frac{KD(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

and

$$\frac{KD^*(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

are identical except perhaps on a set of measure zero. This contradiction proves the theorem.

THEOREM 4. *A necessary and sufficient condition that the power of D be a function of λ only, is that, with the usual exception of a set of measure zero, $D(x_1, \cdots, x_k, u)$ be a function only of*

$$\frac{\sum_1^k x_i^2}{u^2}.$$

The proof of this theorem is not essentially different from that of the preceding theorem, and we shall therefore sketch it only briefly. Let Z be a transformation on $(x_1, \cdots, x_k, u) = (x, u)$ which consists of a rotation of the vector x, followed by a multiplication of u and the components of x by a positive constant c. If $D(x, u)$ is not a function of $\Sigma_1^k x_i^2/u^2$ alone, then, just as before[3], we can use some

[2] See footnote 1.

[3] This statement implies that a function of $x_1, \cdots, x_k, u$, which is invariant to within sets of measure zero under all transformations Z (the exceptional set may depend on the

transformation Z to give us a function $D^*(x, u)$ such that

$$D(x, u) \gtrless D^*(x, u)$$

on a set of positive measure, while

$$ED(x, u) = ED^*(x, u)$$

identically in η, σ. This yields a contradiction in the usual manner and proves the necessity of the condition.

To prove sufficiency, write $D(x, u) = \nu(\Sigma x_i^2/u^2) = \nu(v)$. Let $\gamma(v, \eta, \sigma)$ be the density function of v. Then

$$P\{D \mid \eta, \sigma\} = \int_0^\infty \nu(v)\gamma(v, \eta, \sigma)\, dv.$$

By hypothesis, $\nu(v)$ is a function only of v. We know [9, p. 140, eq. 101] that $\gamma(v, \eta, \sigma)$ is a function only of v and λ. Hence $P\{D \mid \eta, \sigma\}$ is a function only of λ. This completes the proof of the theorem.

THEOREM 5. *Among all tests of the general linear hypothesis which have the properties described in the conclusions of Theorems* 1 *and* 3, *the classical analysis of variance test is the most powerful.*

We shall omit the proof of this theorem, which is very similar to that of the more difficult Theorem 9 below.

Theorem 4 above shows that there exist regions D which satisfy the conclusions of Theorems 1 and 3 and such that $P\{D \mid \eta, \sigma\}$ is not a function of λ alone. It follows that the content of Theorem 5 is greater than that of Hsu's theorem (Corollary to Theorem 2).

It is instructive to note that Hsu's theorem follows almost immediately from Theorem 4 and the form of $\gamma(v, \lambda)$. For let λ be fixed but arbitrary. One verifies immediately from the form of $\gamma(v, \lambda)$ that

$$\frac{\gamma(v, \lambda)}{\gamma(v, 0)}$$

is, for fixed λ, a monotonically increasing function of v. This, by Neyman's lemma, immediately proves Hsu's result.

4. The multiple correlation coefficient. We shall now apply our methods to a multivariate test. For typographic ease we shall conduct the discussion for the

---

transformation), is a function of $\frac{\Sigma x_i^2}{u^2}$, except on a set of measure zero This statement would be completely trivial were it not for the exceptional sets; in any case it must be well known to set theorists The author constructed an unnecessarily long proof of it, and believes that a more expeditious proof can be constructed using the ideas of [11, page 91, Theorem 11 1, and page 318, p. 7]. Professor C. M. Stein of the University of California has informed the author that this result is a special case of one established by himself and G H Hunt in a forthcoming paper. For these reasons the proof is omitted. (See also [13, page 27, Lemma 9 1])

case of three variates, but the reader will observe that the procedure is really perfectly general.

The chance variables $\{Y_{ij}\}$, $i = 1, 2, 3$, $j = 1, \cdots, n$, have the density function

$$g(B) = (2\pi)^{(-3n)/2} (|B|)^{n/2} \exp\left\{-\tfrac{1}{2} \sum_{j=1}^{n} \sum_{i,l=1}^{3} b_{il} y_{ij} y_{lj}\right\} \tag{4.1}$$

where 1) $B = \{b_{il}\}$ is a positive definite (symmetric) 3×3 matrix, 2) y_{ij} is the value assumed by Y_{ij}. The null hypothesis H_0 asserts that a given multiple correlation coefficient is zero, say that of Y_1 with Y_2 and Y_3, i.e.,

$$b_{12} = b_{21} = b_{13} = b_{31} = 0 \tag{4.2}$$

The test is to be made on the level of significance α, i.e., if B_0 is any matrix which satisfies (4.2), and if G is a critical region for testing H_0, then

$$P\{G \mid B_0\} = \alpha \tag{4.3}$$

where the symbol in the left member means the probability of G according to $g(B_0)$.

Write

$$ns_{ij} = \sum_{k=1}^{n} y_{ik} y_{jk}$$

$$S = \begin{Vmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{Vmatrix}.$$

Let $M(c_{11}, C)$ be the manifold in the $3n$-space of

$$y_{11}, \cdots, y_{ik}, \cdots, y_{3n}$$

where $s_{11} = c_{11}$, $S = C$. First we prove the following.

THEOREM 6. *Any region G which satisfies* (4.3) *must have the property that the fraction of the area of $M(c_{11}, C)$ which lies in G is α, for any positive c_{11} and any positive definite 2×2 matrix $C = \{c_{ij}\}$.* (We remind the reader that exceptional sets of measure zero are not precluded).

PROOF. Let $\psi(c_{11}, C)$ be the fraction of the area of $M(c_{11}, C)$ in G Recall equation (4.3) and the fact that s_{11}, s_{22}, s_{23}, s_{33} are sufficient statistics for the elements of B_0. On the manifold $M(c_{11}, C)$ the conditional density is uniform. Employing Wishart's distribution [6] we conclude that

$$K' \int \psi(s_{11}, S) \mid B_0 \mid N \mid S \mid^{(n-3)/2} s_{11}^{(n-2)/2} \cdot \exp\left[-\frac{n}{2} \{b_{11} s_{11} + b_{22} s_{22} + 2b_{23} s_{23} + b_{33} s_{33}\}\right] ds_{11}\, ds_{22}\, ds_{23}\, ds_{33} \equiv \alpha \tag{4.4}$$

where K' is a suitable constant which need not concern us Here the symbol

"$\equiv$" means identically in b_{11}, b_{22}, b_{23}, b_{33}, provided only that $b_{11} > 0$, $b_{22} > 0$ $b_{22}b_{33} - b_{23}^2 > 0$. Of course s_{11} is distributed independently of s_{22}, s_{23}, s_{33}. Proceeding as in section 2, we can, by differentiation with respect to the b's, obtain all the moments of the s_{ij}'s. Now let the b's take any admissible constant values. The moments of the s_{ij}'s are then seen to satisfy the criterion of Cramér and Wold [7, Th. 2], and consequently essentially uniquely determine the distribution of the s_{ij}. The desired conclusion follows as before.

The six parameters which uniquely determine the trivariate normal distribution (of Y_1, Y_2, Y_3) with zero means may be taken to be the following:

1) The covariance matrix $\{\sigma_{ij}\}$, $i, j = 2, 3$, of Y_2 and Y_3.

2) The partial regression coefficients β_2, β_3, of Y_1 on Y_2 and Y_3. These are defined as follows: Let $E(Y_1 \mid Y_2 = y_2, Y_3 = y_3)$ denote the conditional expected value of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$. Then

$$E(Y_1 \mid Y_2 = y_2, Y_3 = y_3) = \beta_2 y_2 + \beta_3 y_3 .$$

3) The conditional variance ω^2 of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$.

The population multiple correlation coefficient $\bar{R}$ of Y_1 with Y_2 and Y_3 is then defined by

$$\frac{\bar{R}^2\omega^2}{(1 - \bar{R}^2)} = \beta_2^2\sigma_{22} + 2\beta_2\beta_3\sigma_{23} + \beta_3^2\sigma_{33} .$$

The six parameters above may be chosen arbitrarily, provided only that $\{\sigma_{ij}\}$ is positive definite. $\bar{R}$ and ω are, by definition, non-negative.

Let y_i be the column vector $y_{i1}, \cdots, y_{in}$; let y_i' be its transpose, and let y denote the point $y_{11}, y_{12}, \cdots, y_{1n}, y_{21}, \cdots, y_{3n}$ in $3n$-space. Let $z(y) = z(y_1, y_2, y_3)$ be the component of y_1 in the plane of y_2 and y_3; let $r = |z(y)|$ and θ the angle between z and y_2, measured positively say in the direction of y_3. Finally let h be the absolute value of the vector $y_1 - z(y_1, y_2, y_3)$.

We intend now to investigate the set theoretic structure of tests whose power is a function only of $\bar{R}$, and for this purpose prove the following:

THEOREM 7. *Let H be a region whose power is a function only of $\bar{R}$. Let $V(h, r, \theta, s_{22}, s_{23}, s_{33})$ be the fraction of the "volume" of the manifold on which $h, r, \theta, s_{22}, s_{23}, s_{33}$ are fixed which is contained in H. With the usual exception of a set of measure zero, for fixed $h, r, s_{22}, s_{23}, s_{33}$, the quantity V above is constant for all θ.*

Later, after this theorem is proved, we shall write V without exhibiting θ. This procedure is justified by Theorem 7.

PROOF. Suppose the theorem false, and proceed as in Theorem 3. A suitable[4] rotation of the radius vector $z(y)$ implies an orthogonal transformation T on the generic point y which leaves h, r, s_{22}, s_{23}, and s_{33} unaltered, and takes the region H into a region H^* such that H and H^* differ on a set of positive measure. T leaves R invariant, hence leaves invariant $\bar{R}$ which uniquely determines the distribution

[4] See footnote 1.

of R Hence an argument almost the same as that which led us to (3.5) yields the conclusion that the power of H and the power of H^* are equal, identically in B. Proceeding as in Theorem 3, we obtain two essentially different density functions in h, r, θ, s_{22}, s_{23}, s_{33}, whose integrals over the entire space are identical in the elements of B. From these functions we obtain two different density functions in $s_{ij}(i, j = 1, 2, 3)$, with identical moments (obtained by differentiation with respect to the elements of B). The rest of the proof is essentially no different from that of Theorem 3.

THEOREM 8. *In order that the power of H be a function of $\bar{R}$ alone, it is necessary and sufficient that, with the usual exception of a set of measure zero, $V(h, r, s_{22}, s_{23}, s_{33})$ be a function only of h/r* (i.e., *of R*).

The proof of this theorem is essentially the same as the proof of Theorem 4 The place of the transformation Z is taken by one which consists of any linear transformation on the vectors y_2 and y_3, the addition of a constant angle to θ (rotation of $z(y)$), and multiplication of the vector y_1 by a positive scalar c. This transformation leaves $\bar{R}$ invariant. In the proof of sufficiency we use the distribution of R (see, for example, [10, p. 384, equation (15 55)]) The remainder of the proof is essentially the same as that of Theorem 4

THEOREM 9. *Among all tests H which have the properties described in the conclusions of Theorems* 6 *and* 7, *the classical test based on R is the most powerful*

As a corollary to this theorem we have the following result due to Simaika [3]: Of all tests H whose power is a function of $\bar{R}$ only, the classical test based on R is the most powerful.

Simaika's result also follows easily from Theorem 8 and the density function of R in the same manner that Hsu's result followed from Theorem 4 and the density function of v.

In the course of the proof of Theorem 9, the various symbols W, with or without subscripts, will denote suitable functions of the variables exhibited, and the various symbols k, with or without subscripts, will denote suitable constants.

We have that

$$P\{H \mid B\} = \int_H (2\pi)^{(-3n)/2} \mid B \mid^{n/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n} y_i' B y_i\right\} dy_{11} \cdots dy_{3n}$$

$$= \int_H (2\pi\omega^2)^{(-n)/2} \exp\left[-\frac{1}{2\omega^2}\{y_1 - (\beta_2 y_2 + \beta_3 y_3)\}^2\right] \cdot$$

$$\cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n} = (2\pi\omega^2)^{(-n)/2} \int_H \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)' z\right\} \cdot \tag{4.5}$$

$$\exp\left[-\frac{1}{2\omega^2}\{y_1^2 + \beta_2^2 s_{22} + 2\beta_2\beta_3 s_{23} + \beta_3^2 s_{33}\}\right] \cdot$$

$$\cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n}.$$

Now $(\beta_2 y_2 + \beta_3 y_3)' z$ is a function only of β_2, β_3, s_{22}, s_{23}, s_{33}, r, and θ. Also

$h^2 + r^2 = s_{11} = y_1^2$. Thus

$$P\{H \mid B\} = \int V(h, r, s_{22}, s_{23}, s_{33}) W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\})$$

$$\cdot \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta\, dh\, dr\, ds_{22}\, ds_{23}\, ds_{33} = \int V(h, r\, s_{22}, s_{23}, s_{33})$$

$$\cdot W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\})(4hr)^{-1} \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} \tag{4.6}$$

$$\cdot d\theta\, dh^2\, dr^2\, ds_{22}\, ds_{23}\, ds_{33} = \int V(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33})$$

$$\cdot W_2(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\}) \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\}$$

$$\cdot d\theta\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}.$$

Integrating with respect to θ and designating

$$W_2 \int \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta$$

by $W(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\})$ we observe that just as in (2.4), W is monotonically increasing in r (all other variables fixed). Thus we have

$$P\{H \mid B\} = \int VW\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}. \tag{4.7}$$

In constructing H only the function V is at our disposal, and this subject to the limitations imposed by the conclusions of Theorems 6 and 7 and the fact that $h^2 + r^2 = y_1^2 = s_{11}$. The function W is not within our control at all. With y_1^2, s_{22}, s_{23}, s_{33} fixed, W is monotonically increasing with r. To maximize the power it is therefore best to distribute the "mass" so that V is as large as possible for large values of r and hence of R. This implies the classical test and proves the theorem.

5. Stringency of the classical tests. Wald [8] calls a test T_1 "most stringent" if the following is true: Let $\{T\}$ be the totality of tests. Let θ be the generic point in the parameter space, and $P\{T \mid \theta\}$ be the power of T at the point θ. Let T_2 be any test other than T_1. Then

$$\sup_\theta [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_1 \mid \theta\}] \leq \sup_\theta [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_2 \mid \theta\}].$$

Of course, we have omitted to specify the totality $\{T\}$. One can admit all tests whose size $\leq \alpha$, a given constant between 0 and 1, or restrict one's self to tests whose size is exactly α. We shall do the latter.

Under these circumstances we shall prove that the classical test of a linear hypothesis is most stringent. Our proof will occupy but a few lines, and is an easy

consequence of the structure of the classical tests as described in the lemma of section 2 The result itself is a special case of an unpublished theorem due to G. H. Hunt and C M Stein, and all priority on this result is theirs

Return then to the notation of section 2. Let σ be fixed at any arbitrary positive value, and the surface

$$\eta^2 = c_0^2$$

be that one on which

$$\omega_1(\eta) = \sup_{\{T\}} P\{T \mid \eta\} - P\{L_1 \mid \eta\}$$

is a maximum, where L_1 is the classical test of the linear hypothesis. It is clear that this maximum is actually achieved, and that $\omega_1(\eta)$ is a constant on the surface $\eta^2 = c_0^2$. Let L_2 be any other test (of size α), and $\omega_2(\eta)$ be the corresponding function for L_2. We have only to show that on the surface $\eta^2 = c_0^2$ we cannot have everywhere $\omega_2(\eta) < \omega_1(\eta)$, and our proof is complete. If everywhere on the surface $\eta^2 = c_0^2$ we had $\omega_2(\eta) < \omega_1(\eta)$, we would have, also on the same surface, $P\{L_2 \mid \eta\} > P\{L_1 \mid \eta\}$. This would, however, violate Wald's Theorem 2 (section 2) and proves the desired result

REFERENCES

[1] P L Hsu, "Notes on Hotelling's generalized T," *Annals of Math Stat.*, Vol. 9 (1938) p. 231.

[2] P. L Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol 32 (1941), p. 62

[3] J B Simaika, "On an optimum property of two important statistical tests," *Biometrika*, Vol 32 (1941), p 70.

[4] A Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p 434.

[5] J A Shohat and J D Tamarkin, *The Problem of Moments*, The American Mathematical Society, New York, 1943.

[6] John Wishart, "The generalized product moment distribution, etc ," *Biometrika*, Vol 20A (1928), p 32.

[7] H. Cramér and H. Wold, "Some theorems on distribution functions," *Lond Math. Soc. Jour* , Vol. 11 (1936)

[8] A Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Am. Math Soc. Trans* , Vol 54 (1943), p 426

[9] P. C Tang, "The power function of the analysis of variance etc ," *Stat. Res. Memoirs*, Vol. 2 (1938) (University of London), p. 126.

[10] M G. Kendall, *The Advanced Theory of Statistics*, Vol 1, Charles Griffin and Company, London, 1945.

[11] S Saks, *Theory of the Integral*, (Second Edition), G E Stechert and Company, New York, 1937

[12] J. Neyman and E. S Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. London Phil. Trans* , Ser A, Vol. 231 (1933), pp. 289-337.

[13] Eberhard Hopf, *Ergodentheorie*, Chelsea, New York, 1948

APPLICATION OF THE METHOD OF MIXTURES TO QUADRATIC FORMS IN NORMAL VARIATES

By Herbert Robbins and E. J. G. Pitman

Institute of Statistics, University of North Carolina

1. Summary. The method of mixtures, explained in Section 2, is applied to derive the distribution functions of a positive quadratic form in normal variates and of the ratio of two independent forms of this type.

2. The method of mixtures. If

$$F_0(x), \qquad F_1(x), \tag{1}$$

is any sequence of distribution functions, and if

$$c_0, c_1, \cdots \tag{2}$$

is any sequence of constants such that

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1 \tag{3}$$

(all summations will be from 0 to ∞ unless otherwise noted), then the function

$$F(x) = \Sigma c_j F_j(x) \tag{4}$$

is called a *mixture* of the sequence (1).

It is sometimes helpful to interpret $F(x)$ in the following manner. Let $J, X_0, X_1, \cdots$ be variates such that J has the distribution $P[J = j] = c_j$ $(j = 0, 1, \cdots)$ and such that X_j has the distribution function $F_j(x)$. Let X be a variate such that the conditional distribution function of X given $J = j$ is $F_j(x)$. Then the distribution function of X is

$$P[X \leq x] = \Sigma P[J = j] \cdot P[X \leq x \mid J = j] = \Sigma c_j F_j(x) = F(x).$$

This interpretation of $F(x)$ will, however, not be involved in the present paper.

The following statements are proved in [1]. If $x = (x_1, \cdots, x_n)$ is a vector variable the function $F(x)$ defined by (4) is a distribution function, and for any Borel set S,

$$\int_S dF(x) = \Sigma c_j \int_S dF_j(x). \tag{5}$$

More generally, if $g(x)$ is any Borel measurable function then

$$\int_{-\infty}^{\infty} g(x)\, dF(x) = \Sigma c_j \int_{-\infty}^{\infty} g(x)\, dF_j(x) \tag{6}$$

whenever the left hand side of (6) exists. In particular, the characteristic function

$\varphi(t)$ corresponding to $F(x)$ is

$$\varphi(t) = \Sigma c_j \varphi_j(t), \tag{7}$$

where $\varphi_j(t)$ is the characteristic function corresponding to $F_j(x)$.

If each $F_j(x)$ has a derivative $f_j(x)$ then $F(x)$ has a derivative $f(x)$ given by

$$f(x) = \Sigma c_j f_j(x), \tag{8}$$

provided that this series converges uniformly in some interval including x. Conversely, if (8) is the relation between the frequency functions and if the series is uniformly convergent in every finite interval, then the relation between the distribution functions is given by (4). In practice we deduce (4) from (8), or, using the uniqueness theorem for characteristic functions, from (7).

As regards computation, we observe that for any integers $0 \leq p_1 \leq p_2$ and for any x it follows from (3) and (4) that

$$\begin{aligned} 0 \leq F(x) - \sum_{p_1}^{p_2} c_j F_j(x) &= \sum_{0}^{p_1-1} c_j F_j(x) + \sum_{p_2+1}^{\infty} c_j F_j(x) \\ &\leq \sup_{j<p_1} \{F_j(x)\} \cdot \left(\sum_{0}^{p_1-1} c_j \right) + \sup_{j>p_2} \{F_j(x)\} \cdot \left(1 - \sum_{0}^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j \right) \leq 1 - \sum_{p_1}^{p_2} c_j. \end{aligned} \tag{9}$$

The existence of these upper bounds (the last a uniform one) for the error term when the series (4) is replaced by a finite sum shows that series expansions of the mixture type (4) are especially well adapted to computational work.

For some purposes it is useful to consider series expansions of the type (4) where the c_j may be of both signs and where the series Σc_j may diverge. Both parts of (3) will, however, be satisfied in the cases considered here.

If U, V are independent variates with respective distribution functions $F(x)$, $G(x)$ we shall denote the distribution function of any Borel measurable function $H(U, V)$ by

$$H(U, V)\ (F(x), G(x)).$$

Now if $F(x)$, $G(x)$ are both mixtures,

$$F(x) = \Sigma c_j F_j(x), \qquad G(x) = \Sigma d_k G_k(x),$$

then by (5),

$$\begin{aligned} P[H(U, V) \leq x] &= \iint_{\{H(u,v) \leq x\}} dF(u)\, dG(v) \\ &= \Sigma\Sigma c_j\, d_k \iint_{\{H(u,v) \leq x\}} dF_j(u)\, dG_k(v), \end{aligned}$$

so that

$$H(U, V)(\Sigma c_j F_j(x), \Sigma d_k G_k(x)) = \Sigma\Sigma c_j\, d_k\, H(u, v)(F_j(x), G_k(x)). \tag{10}$$

As an application of the principles set forth in this section we shall express as series of the mixture type (4) the distribution functions of any positive quadratic form in normal variates and of the ratio of any two independent forms of this type. Special cases of the problem have been dealt with by Tang [2], Hsu [3], and many others, but the method of mixtures permits a unified and simple treatment of the general case.

3. Distribution of a positive quadratic form. We shall denote by $F_n(x)$ the chi-square distribution function with $n > 0$ degrees of freedom,

$$F_n(x) = \frac{1}{2^{\frac{1}{2}n} \cdot \Gamma(\frac{1}{2}n)} \int_0^x u^{\frac{1}{2}n-1} \cdot e^{-\frac{1}{2}u} \cdot du \qquad (x > 0), \tag{11}$$

$$= 0 \qquad (x \leq 0)$$

The corresponding characteristic function is

$$\varphi_n(t) = \int_0^\infty e^{ixt}\, dF_n(x) = (1 - 2it)^{-\frac{1}{2}n} = w^{\frac{1}{2}n}, \tag{12}$$

where we have set $w = (1 - 2it)^{-1}$. We shall denote by χ_n^2 any variate with the distribution function (11).

Let a be any constant such that $a > 0$. The characteristic function of the variate $a \cdot \chi_n^2$ is

$$(1 - 2iat)^{-\frac{1}{2}n} = [a(1 - 2it) - (a - 1)]^{-\frac{1}{2}n} = a^{-\frac{1}{2}n} \cdot w^{\frac{1}{2}n} \cdot \left(1 - \left(1 - \frac{1}{a}\right)w\right)^{-\frac{1}{2}n}. \tag{13}$$

By the binomial theorem we have for any $a > 0$,

$$a^{-\frac{1}{2}n}\left[1 - \left(1 - \frac{1}{a}\right)z\right]^{-\frac{1}{2}n} = \Sigma c_j z^j \qquad \left(|z| < \left|1 - \frac{1}{a}\right|^{-1}\right), \tag{14}$$

where

$$c_j = a^{-\frac{1}{2}n} \cdot \frac{\frac{1}{2}n(\frac{1}{2}n + 1)\cdots(\frac{1}{2}n + j - 1)}{j!} \cdot \left(1 - \frac{1}{a}\right)^j \qquad (j = 0, 1, \cdots). \tag{15}$$

For $a \geq 1$ we see from (15) that all the c_j are non-negative. Likewise for $a > \frac{1}{2}$ (and hence *a fortiori* for $a \geq 1$) we have $|1 - 1/a|^{-1} > 1$ so that (14) holds for all $|z| \leq 1$; setting $z = 1$ it follows that the sum of all the c_j is equal to 1. Hence for $a \geq 1$,

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1.$$

Since $|w| = |1 - 2it|^{-1} \leq 1$ for all real t it follows from (13) and (14) that for $a \geq 1$,

$$(1 - 2iat)^{-\frac{1}{2}n} = \Sigma c_j\, w^{\frac{1}{2}n+j} = \Sigma c_j (1 - 2it)^{-\frac{1}{2}n-j}$$

$$= \Sigma c_j\, \varphi_{n+2j}(t). \tag{16}$$

Hence for $a \geq 1$ the distribution function $F_n(x/a)$ of the variate $a \cdot \chi_n^2$ is a mixture of χ^2 distribution functions,

$$F_n(x/a) = \Sigma c_j F_{n+2j}(x), \tag{17}$$

where the c_j, determined by the identity (14), are the probabilities of a negative binomial distribution.

It may, in fact be proved by a direct analysis, which we omit here, that (17) holds for any $a > 0$. However, if $a < 1$ then the c_j will be of alternating sign, and if $a \leq \frac{1}{2}$ then the series Σc_j will diverge This shows incidentally that a relation of the form (4) can hold even though the series Σc_j diverges and hence the corresponding relation (7) does not hold for $t = 0$.

THEOREM 1. *Let*

$$X = a(\chi_m^2 + a_1 \chi_{m_1}^2 + \cdots + a_r \chi_{m_r}^2),$$

where the chi-square variates are independent and $a, a_1, \cdots, a_r$ *are positive constants such that*

$$a_i \geq 1 \qquad (i = 1, \cdots, r).$$

Define constants c_j *by the identity*[1]

$$\prod_{i=1}^{r} \left\{ a_i^{-\frac{1}{2}m_i} \left[1 - \left(1 - \frac{1}{a_i} \right) z \right]^{-\frac{1}{2}m_i} \right\} = \Sigma c_j z^j \qquad (|z| \leq 1); \tag{18}$$

then obviously

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1.$$

Let

$$M = m + m_1 + \cdots + m_r ;$$

then for every x,

$$P[X \leq x] = \Sigma c_j \cdot F_{M+2j}(x/a). \tag{19}$$

For any integers $0 \leq p_1 \leq p_2$ *and every* x,

$$\begin{aligned} 0 \leq P[X \leq x] &- \sum_{p_1}^{p_2} c_j F_{M+2j}(x/a) \\ &\leq F_M(x/a) \cdot \left(\sum_{0}^{p_1-1} c_j \right) + F_{M+2p_2+2}(x/a) \cdot \left(1 - \sum_{0}^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j \right) \\ &\leq 1 - \sum_{p_1}^{p_2} c_j . \end{aligned} \tag{20}$$

PROOF. The characteristic function of X/a is, by (13) and (18),

$$\varphi(t) = w^{\frac{1}{2}M} \cdot \prod_{i=1}^{r} \left\{ a_i^{-\frac{1}{2}m_i} \left[1 - \left(1 - \frac{1}{a_i} \right) w \right]^{-\frac{1}{2}m_i} \right\} = \Sigma c_j w^{\frac{1}{2}M+j} = \Sigma c_j \varphi_{M+2j}(t)$$

[1] If $r = 0$ we regard the left hand side of (18) as having the value 1.

Hence for any y,

$$P[X/a \leq y] = \Sigma c_j F_{m+2j}(y),$$

whence (19) follows on setting $x = ay$. Finally, since $F(x)$ is a decreasing function of n for fixed x, (20) follows from (9).

It should be observed that the coefficients c_j determined by (18) can be written explicitly as the multiple Cauchy products

$$c_j = \sum_{i_1+\cdots+i_r=j} \{c_{1,i_1} \cdots c_{r,i_r}\},$$

where

$$c_{i,j} = a_i^{-\frac{1}{2}m_i} \cdot \frac{\frac{1}{2}m_i(\frac{1}{2}m_i + 1) \cdots (\frac{1}{2}m_i + j - 1)}{j!} \cdot \left(1 - \frac{1}{a_i}\right)^j$$

$$(i = 1, \cdots, r; j = 0, 1, \cdots).$$

The c_j may be computed stepwise by the relations

$$c_j^{(1)} = c_{1,j},$$

$$c_j^{(s)} = \sum_{i=0}^{j} \{c_{j-i}^{(s-1)} \cdot c_{s,i}\} \qquad (s = 2, \cdots, r),$$

$$c_j^{(r)} = c_j.$$

4. Distribution of a ratio. The ratio χ_m^2/χ_n^2 of two independent chi-square variates has the distribution function

$$(21) \qquad F_{m,n}(x) = \frac{\Gamma(\frac{1}{2}(m+n))}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} \int_0^x u^{\frac{1}{2}m-1}(1+u)^{-\frac{1}{2}(m+n)}\, du \qquad (x \geq 0),$$

$$= 0 \qquad (x < 0).$$

In computational work we can use the tables of the Beta distribution function

$$I_x(r, s) = \frac{\Gamma(r+s)}{\Gamma(r) \cdot \Gamma(s)} \int_0^x u^{r-1} \cdot (1-u)^{s-1} \cdot du \qquad (0 < x < 1),$$

$$= 0\ (x \leq 0), \qquad 1\ (x \geq 1),$$

together with the identity

$$F_{m,n}(x) = I_{x/(1+x)}(\tfrac{1}{2}m, \tfrac{1}{2}n).$$

THEOREM 2. *Let*

$$(22) \qquad X = \frac{a \cdot (\chi_m^2 + a_1\chi_{m_1}^2 + \cdots + a_r\chi_{m_r}^2)}{\chi_n^2 + b_1\chi_{n_1}^2 + \cdots + b_s\chi_{n_s}^2},$$

where the χ^2 variates are independent and a, $a_1, \cdots, a_r$, $b_1, \cdots, b_s$ are positive

constants such that

$$a_i \geq 1, \qquad b_j \geq 1 \qquad (i = 1, \cdots, r; j = 1, \cdots, s).$$

Define constants c_j, d_k *by the identities*

$$\prod_{i=1}^{r}\left\{a_i^{-\frac{1}{2}m_i}\cdot\left[1-\left(1-\frac{1}{a_i}\right)z\right]^{-\frac{1}{2}m_i}\right\} = \Sigma c_j z^j, \qquad (|z| \leq 1)$$
$$\prod_{i=1}^{s}\left\{b_i^{-\frac{1}{2}n_i}\cdot\left[1-\left(1-\frac{1}{b_i}\right)z\right]^{-\frac{1}{2}n_i}\right\} = \Sigma\, d_k z^k;$$

then

$$c_j \geq 0, \quad \Sigma c_j = 1, \qquad d_k \geq 0, \quad \Sigma d_k = 1.$$

Let

$$M = m + m_1 + \cdots + m_r, \qquad N = n + n_1 + \cdots + n_s;$$

then for every x,

$$P[X \leq x] = \Sigma\Sigma c_j\, d_k \cdot F_{M+2j,N+2k}(x/a),$$

and for any integers $0 \leq p_1 \leq p_2$, $0 \leq q_1 \leq q_2$ *and every* x,

$$0 \leq P[X \leq x] - \sum_{p_1}^{p_2}\sum_{q_1}^{q_2} c_j\, d_k \cdot F_{M+2j,N+2k}(x/a) \leq \left(1 - \sum_{p_1}^{p_2} c_j\right)\cdot\left(1 - \sum_{q_1}^{q_2} d_k\right).$$

PROOF. Let U, V denote respectively numerator and denominator of (22). From Theorem 1,

$$P[U \leq x] = \Sigma c_j F_{M+2j}(x/a),$$
$$P[V \leq x] = \Sigma\, d_k F_{N+2k}(x).$$

Hence by (10), for every x,

$$P[X \leq x] = P[U/V \leq x] = \Sigma\Sigma c_j\, d_k \cdot F_{M+2j,N+2k}(x/a).$$

The rest of the theorem is obvious.

COROLLARY. *Let*

$$X = \frac{\chi_M^2}{a\chi_r^2 + b\chi_s^2},$$

where the χ^2 *variates are independent and*

$$0 < a \leq b.$$

Define

$$\alpha = a/b, \qquad N = r + s,$$

$$c_j = \alpha^{\frac{1}{2}s} \cdot \frac{\frac{1}{2}s(\frac{1}{2}s+1)\cdots(\frac{1}{2}s+j-1)}{j!} \cdot (1-\alpha)^r \qquad (j = 0, 1, \cdots);$$

then

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1,$$

and for every x,

$$P[X \leq x] = \Sigma c_j F_{M,N+2j}(ax).$$

For any integers $0 \leq p_1 \leq p_2$ *and every* x,

$$0 \leq p[X > x] - \sum_{p_1}^{p_2} c_j[1 - F_{M,N+2j}(ax)]$$

$$\leq [1 - F_{M,N}(ax)] \cdot \left(\sum_{0}^{p_1-1} c_j\right) + [1 - F_{M,N+2p_2+2}(ax)] \cdot \left(1 - \sum_{0}^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j\right) \leq 1 - \sum_{p_1}^{p_2} c_j. \tag{23}$$

PROOF. Except for (23) this is a special case of Theorem 2. To prove (23) we observe that

$$P[X > x] = 1 - P[X \leq x] = \Sigma c_j[1 - F_{M,N+2j}(ax)],$$

and since for fixed m and x, $F_{m,n}(x)$ is an increasing function of n, (23) follows in the same way as (9).

5. The non-central case. Let Y be normal (0, 1) and let $X = (Y + d)^2$, where d is any constant. The frequency function of X is, for $x > 0$,

$$f(x) = (2\pi x)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(d^2+x)} \cdot (e^{dx^{\frac{1}{2}}} + e^{-dx^{\frac{1}{2}}})/2.$$

By expanding the last factor into a power series it is easily seen that

$$f(x) = \Sigma p_j \cdot f_{1+2j}(x), \tag{24}$$

where $f_n(x) = F_n'(x)$ is the chi-square frequency function with n degrees of freedom and where

$$p_j = e^{-\frac{1}{2}d^2} \cdot (\tfrac{1}{2}d^2)^j/j! \qquad (j = 0, 1, \cdots).$$

Since the identity

$$e^{-\frac{1}{2}d^2(1-z)} = \Sigma p_j z^j \qquad \text{(all } z\text{)} \tag{25}$$

holds, it follows that

$$p_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma p_j = 1.$$

The series (24) is uniformly convergent in every finite interval, so that we can write the distribution function $F(x)$ and characteristic function $\varphi(t)$ of X in the forms

$$F(x) = \Sigma p_j \cdot F_{1+2j}(x),$$
$$\varphi(t) = \Sigma p_j \cdot \varphi_{1+2j}(t) = w^{\frac{1}{2}} \cdot e^{-\frac{1}{2}d^2(1-w)},$$

where again we have set $w = (1 - 2it)^{-1}$.

Now let $Y_1, \cdots, Y_n$ be independent and normal (0, 1) variates and let

$$X = (Y_1 + d_1)^2 + \cdots + (Y_n + d_n)^2, \tag{26}$$

where the d_i are constants such that

$$d_1^2 + \cdots + d_n^2 = d^2.$$

The characteristic function of X is then

$$\varphi(t) = w^{\frac{1}{2}n} e^{-\frac{1}{2}d^2(1-w)} = \Sigma p_j w^{\frac{1}{2}n+j} = \Sigma p_j \varphi_{n+2j}(t),$$

and hence the distribution function $F(x)$ of X is again a mixture of χ^2 distribution functions,

$$F(x) = \Sigma p_j \cdot F_{n+2j}(x), \tag{27}$$

where the p_j, determined by the identity (25), are the probabilities of a Poisson distribution with parameter $\lambda = \frac{1}{2}d^2$. We shall denote the non-central chi-square variate (26) by $\chi'^2_{n,d}$.

We can now generalize Theorems 1 and 2 in a straightforward manner to cover non-central chi-square variates. We shall state only the generalization of the Corollary of Theorem 2 to the case in which the numerator is non-central.

THEOREM 3. *Let*

$$X = \frac{\chi'^2_{M,d}}{a\chi_r^2 + b\chi_s^2},$$

where the χ^2 variates are independent and

$$0 < a \leq b.$$

Define

$$\lambda = \tfrac{1}{2}d^2, \qquad \alpha = a/b, \qquad N = r + s$$
$$p_j = e^{-\lambda} \cdot \lambda^j / j! \qquad (j = 0, 1, \cdots),$$
$$c_k = \alpha^{\frac{1}{2}s} \cdot \frac{\frac{1}{2}s(\frac{1}{2}s + 1)\cdots(\frac{1}{2}s + k - 1)}{k!} \cdot (1 - \alpha)^k \qquad (k = 0, 1, \cdots);$$

then

$$p_j \geq 0, \quad \Sigma p_j = 1, \qquad c_k \geq 0, \quad \Sigma c_k = 1,$$

and for every x,

$$P[X \leq x] = \Sigma\Sigma p_j c_k F_{M+2j,N+2k}(ax).$$

For any integers $0 \leq g_1 \leq g_2$, $0 \leq h_1 \leq h_2$,

$$0 \leq P[X \leq x] - \sum_{g_1}^{g_2}\sum_{h_1}^{h_2} p_j c_k \cdot F_{M+2j,N+2k}(ax) \leq \left(1 - \sum_{g_1}^{g_2} p_j\right) \cdot \left(1 - \sum_{h_1}^{h_2} c_k\right).$$

REFERENCES

[1] HERBERT ROBBINS, "Mixture of distributions," *Annals of Math. Statistics*, Vol. 19 (1948), p. 360

[2] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), p. 126.

[3] P. L. HSU, "Contributions to the theory of 'Student's' t-test as applied to the problem of two samples," *ibid.*, p. 1.

THE JOINT DISTRIBUTION OF SERIAL CORRELATION COEFFICIENTS

By M. H. Quenouille

Rothamsted Experimental Station

1. Summary. An expression for the joint distribution of serial correlation coefficients, circularly defined, has been derived. It has been shown that this distribution possesses properties similar to those already encountered in the distribution of a single serial correlation coefficient, i.e. it is definedby different function forms for various subregions. The distribution thus found is of little use for computational purposes Consequently, approximate forms have been investigated and the suitability of the ordinary partial correlation coefficient for large-sample testing has been inferred.

2. Introduction. Anderson [1] has derived the distribution of the serial correlation coefficient

$$r_l = \frac{\sum_{i=1}^{n} \epsilon_i \epsilon_{i+l} - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n}{\sum_{i=1}^{n} \epsilon_i^2 - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n},$$

where the ϵ_i are normally and independently distributed with mean μ and variance σ^2 and where a circular definition is employed, so that ϵ_{n+i} is defined to be equal to ϵ_i. However, in making a test of any series, we shall usually be faced with a set of serial correlation coefficients, so that we shall require a joint distribution function of $r_1, r_2, \cdots, r_m$ say. This distribution function is derived below by an extension of the method used by Koopmans [2]

It should be noted that Bartlett [3] has shown that for large samples the variances and covariances of the r_l are independent of the distribution of ϵ_i under fairly wide conditions This means that the joint distribution function obtained for normal ϵ_i will often give a good approximation for non-normal ϵ_i and can be used as the basis for any test of the correlogram.

3. Conditions on the r_l. It is easily seen that the r_l cannot take all values from $+1$ to -1 independently. For example, r_2 cannot take a value near -1 if r_l takes a value near $+1$. As a result, there will be certain necessary conditions that the r_l will have to fulfil. It is not difficult to find these conditions, since, if $y_i (i = 1, 2, \cdots, n)$ are any set of variables, then

$$\sum_{j=1}^{n} (\epsilon_{i+j} y_i)^2 = \left(\sum_{i=1}^{n} \epsilon_i^2\right) r_j y_l y_{l+j}, \tag{1}$$

where ϵ_i may or may not be corrected for the mean and the double-suffix summation convention is employed.

Thus, provided $0 < m < n/2$, we will have

$$R_m = \begin{vmatrix} 1 & r_1 & r_2 & \cdots & r_m \\ r_1 & 1 & r_1 & \cdots & r_{m-1} \\ r_2 & r_1 & 1 & \cdots & r_{m-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_m & r_{m-1} & r_{m-2} & \cdots & 1 \end{vmatrix} \geq 0 \tag{2}$$

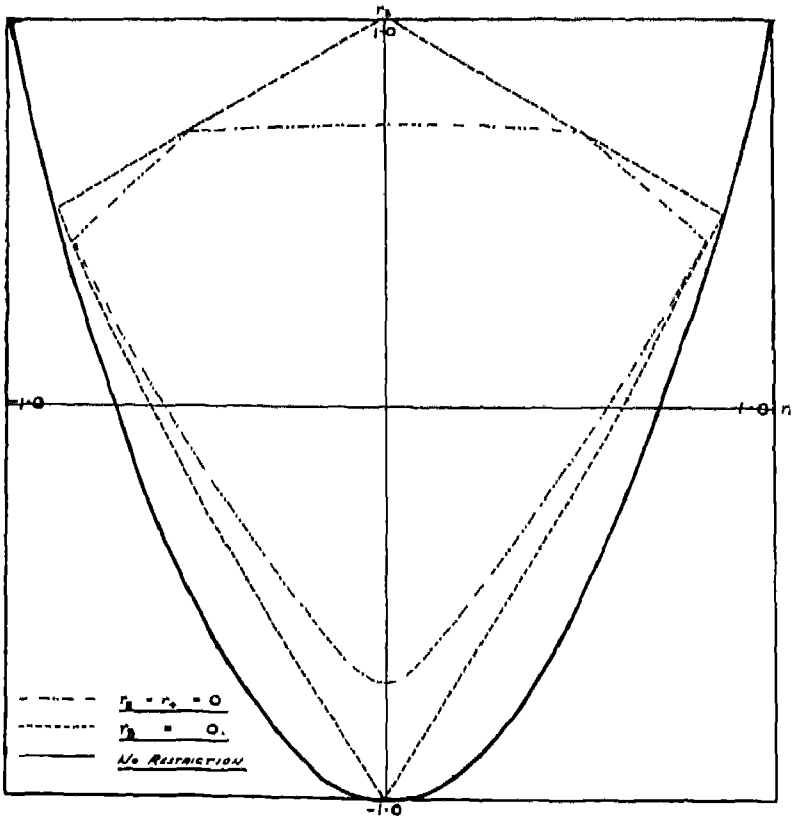

FIG. 1

as a necessary condition that the right-hand side of (1) be positive definite and this expression will impose necessary conditions upon the joint distribution of the r_i.

Fig. 1 gives the limits of possible values of r_1 and r_2 subject to (a) no restriction, (b) $r_3 = 0$, (c) $r_3 = r_4 = 0$.

4. Complex Integration in m Variables. Before finding the joint distribution function of the r_i some introductory remarks on complex integration involving m variables will be necessary.

We can evaluate an integral such as

$$\int_{-\infty}^{\infty} \cdots \int \frac{f(z_1, z_2 \cdots z_m)}{\prod_{j=1}^{m} (z_j - a_j)} \, dz_1 \cdots dz_m$$

where $\mathcal{I}(a_i) = 0$ and $f(z_1, z_2, \cdots, z_m)$ is regular in the region $\mathcal{I}(z_i) \geq 0$, by successive Cauchy integrations, so the integral has a value $(2\pi i)^m f(a_1, \cdots, a_m)$. In the same manner as for Cauchy integration, it will be possible to distort the contours over which we integrate so that we can evaluate

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{m} (z_j - a_j)} \, dz_1 \cdots dz_m,$$

provided that $f(z_1, \cdots, z_m)$ is regular in the region defined by S, and $(a_1, \cdots, a_m)$ is enclosed in this region

More generally, if we have an integral of the form

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{m} (a_{ij} z_i - b_j)} \, dz_1 \cdots dz_m,$$

and we make the transformations $w_j = a_{ij} z_i$ and $b_j = a_{ij} c_i$, i.e. $\boldsymbol{W} = \boldsymbol{AZ}$, $\boldsymbol{C} = \boldsymbol{A}^{-1}\boldsymbol{B}$, it is possible, in the above manner, to evaluate the integral as

$$\pm \frac{(2\pi i)^m}{|A|} f(c_1 \cdots c_m). \tag{3}$$

Suppose we now consider the integral

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{n} (a_{ij} z_i - b_j)} \, dz_1 \cdots dz_m,$$

where $n \geq m$. We may select a set, g_k, of m equations $a_{ij} z_i = b_j$, and let $\boldsymbol{A}_k = [a_{ij}]$, $\boldsymbol{B}_k = [b_j]$, $\boldsymbol{C}_k = \boldsymbol{A}_k^{-1}\boldsymbol{B}_k = [c_{ik}]$. Then, we may carry out the integration as previously, in this case, summing a series of terms for various combinations of m equations out of the possible n. The value of the integral may then be written

$$(2\pi i)^m \sum_{0k} \pm \frac{f(c_{1k}, \cdots c_{nk})}{|A_k| \prod_{l \neq g_k} (a_{l_j} c_{lk} - b_j)}, \tag{4}$$

where the summation occurs over the points $(c_{ik}, c_{2k}, \cdots, c_{nk})$ lying in the region defined by S, and the product term excludes the set of equations g_k. The ambiguity of sign in (3) and (4) arises from the Jacobian $|A_k|^{-1}$, and the sign must be chosen which makes the transformation of $dz_1, \cdots, dz_m$ yield a positive

element. It must be noted that it is possible to obtain several expansions of the form (4) according to the convention that is employed in defining "enclosure" for each of the variables.

5. Integral form for the joint distribution function. We can, without loss of generality, assume $\sigma^2 = 1$. Suppose that

$$p = \sum_{i=1}^{n} \epsilon_i^2 - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n, \qquad q_l = \sum_{i=1}^{n} \epsilon_i \epsilon_{i+l} - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n,$$

where $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independent, so that $r_l = q_l/p$. Then by a consideration of n dimensional space, we can see that p is distributed independently of $r_1, \cdots r_m$ so that their joint distribution can be written $g(p)h(r_1, \cdots, r_m)dp\, dr_1, \cdots, dr_m$. The joint distribution of p and $q_1, \cdots, q_m$ can thus be written

$$f(pq_1 \cdots q_m)\, dp\, dq_1 \cdots dq_m = \frac{g(p)}{p^m} h\left(\frac{q_1}{p}, \cdots \frac{q_m}{p}\right) dp\, dq_1 \cdots dq_m, \tag{5}$$

where it is not difficult to see that

$$g(p) = \frac{p^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p}}{2^{\frac{1}{2}(n-1)} \Gamma\left(\frac{n-1}{2}\right)}. \tag{6}$$

We can now find the joint distribution of p and $q_1, \cdots, q_m$ by inverting the characteristic function of these variables. This is given by

$$\begin{aligned} \frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{-\infty}^{\infty} \cdots \int \exp\left[-\frac{\Sigma\epsilon_j^2}{2} + i(\eta p + \theta_j q_j)\right] d\epsilon_1 \cdots d\epsilon_n, & \\ = \frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{-\infty}^{\infty} \cdots \int \exp\left[-\frac{\epsilon' \Delta \epsilon}{2}\right] d\epsilon_1 \cdots d\epsilon_n, & \\ = 1/|\Delta|^{\frac{1}{2}}, & \end{aligned}$$

where

$$\epsilon' = [\epsilon_1, \epsilon_2, \cdots, \epsilon_n]$$

and

$$\begin{aligned} |\Delta| &= \prod_{l=1}^{n-1} (1 - 2i\eta - 2i\theta_j \kappa_{jl}), \qquad \kappa_{jl} = \cos \frac{2\pi j l}{n}, \\ &= (1 - 2i\eta)^{n-1} \prod_{l=1}^{n-1} (1 - \kappa_j \kappa_{jl}), \qquad \kappa_j = \frac{2i\theta_j}{1 - 2i\eta}, \end{aligned}$$

so that the joint distribution of p and $q_1, \cdots, q_m$ is

$$\begin{aligned} f(p, q_1 \cdots q_m) &= \frac{1}{(2\pi)^{m+1}} \int_{-\infty}^{\infty} \cdots \int \frac{1}{|\Delta|^{\frac{1}{2}}} \exp\{-i(\eta p + \theta_j q_j)\}\, d\eta\, d\theta_1 \cdots d\theta_m \\ &= \frac{1}{(2\pi)^{m+1}} \int_{-\infty}^{\infty} e^{-i\eta p} (1 - 2i\eta)^{-\frac{1}{2}(n-2m-1)} \int_{S_\eta} \frac{1}{|\Delta|^{\frac{1}{2}}} \\ &\qquad \exp\left\{-\frac{(1 - 2i\eta)\kappa_j q_j}{2}\right\} \frac{d\kappa_1 \cdots d\kappa_m}{(2i)^m} d\eta, \end{aligned} \tag{7}$$

where S_η is the region bounded by $\kappa_j = \pm \dfrac{2i\infty}{1 - 2i\eta}$. Now S_η can be replaced by region S enclosing the same set of singularities on the real hyperplane, and S can be chosen independent of η. Thus it will be possible to reverse the order of integration in (7) provided that $\int_{-\infty}^{\infty} |1 - 2i\eta|^{-\frac{1}{2}(n-2m-1)} d\eta$ converges, i.e. provided $n > 2m + 3$. Then since

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(1 - 2i\eta)^{-\frac{1}{2}(n-2m-1)} \exp\{-i\eta(p - \kappa_j q_j)\}\, d\eta$$
$$= \frac{(p - \kappa_j q_j)^{\frac{1}{2}(n-2m-3)}}{2^{\frac{1}{2}(n-2m-1)}\Gamma\left(\dfrac{n-2m-1}{2}\right)} \exp\{-\tfrac{1}{2}(p - \kappa_j q_j)\} \quad \text{for} \quad p \geq \kappa_j q_j,$$
$$= 0 \quad \text{for} \quad p \leq \kappa_j q_j,$$

we get

$$f(p, q_1 \cdots q_m) = \frac{e^{-\frac{1}{2}p}}{2^{\frac{1}{2}(n-1)}(2\pi i)^m \Gamma\left(\dfrac{n-2m-1}{2}\right)} \cdot \int_S \cdots \int \frac{(p - \kappa_j q_j)^{\frac{1}{2}(n-2m-3)}}{\left[\prod_{l=1}^{n-1}(1 - \kappa_j \kappa_l)\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m, \tag{8}$$

where S encloses the same singularities as S_η all of which lie in the region $p \geq \kappa_j q_j$. If we now use (5) and (6) we get

$$h(r_1 \cdots r_m) = \frac{\Gamma\left(\dfrac{n-1}{2}\right)}{\Gamma\left(\dfrac{n-2m-1}{2}\right)} \cdot \frac{1}{(2\pi i)^m} \int_S \cdots \int \frac{(1 - \kappa_j r_j)^{\frac{1}{2}(n-2m-3)}}{\left[\prod_{l=1}^{n-1}(1 - \kappa_j \kappa_l)\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m. \tag{9}$$

In a similar manner, it is possible to derive for $n \geq 2m + 3$ the joint distribution of serial correlation coefficients, $\bar{r}_1, \cdots, \bar{r}_m$, uncorrected for the mean, in the form

$$h(\bar{r}_1 \cdots \bar{r}_m) = \frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{n-2m}{2}\right)} \frac{1}{(2\pi i)^m} \int_S \cdots \int \frac{(1 - \kappa_j \bar{r}_j)^{\frac{1}{2}(n-2m-2)}}{\left[\prod_{l=1}^{n}(1 - \kappa_j \kappa_l)\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m. \tag{10}$$

6. Extension for variables in an autoregressive scheme. Madow [4] has shown how to extend the distribution of the serial correlation coefficient for uncorrelated variables to the case when the variables x_i are connected by a linear Markoff scheme, $x_i = \rho x_{i-1} + \epsilon_i$ with a normal distribution of the error ϵ_i. It is worth

noting that the method used by Madow can be applied to derive the joint distribution of serial correlations of variables x_i, which are connected by a linear autoregressive scheme of order m, or less,

$$a_0 x_i + a_1 x_{i-1} + \cdots + a_m x_{i-m} = \epsilon_i ,$$

where $\epsilon_1 , \cdots , \epsilon_n$ are normally and independently distributed, and $\epsilon_{n+i} = \epsilon_i$.[1] Under these conditions, the expression (9) will be modified by a factor

$$\left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \epsilon_i^2} \right)^{\frac{1}{2}(n-1)} = \frac{1}{(A + 2B_j r_j)^{\frac{1}{2}(n-1)}} , \tag{11}$$

where

$$A = \sum_{k=0}^{m} a_k^2 ,$$

$$B_j = \sum_{k=0}^{m-j} a_k a_{k+j} ,$$

while (10) will be modified by a similar factor with n replacing $n - 1$.

7. Reduction of the distribution function integral. Using the method described in section 4, it is now possible to reduce the integral given in (9), if we observe that $\kappa_{jl} = \kappa_{jn-l}$ and assume n odd. We then have

$$\begin{aligned} h(r_1 \cdots r_m) &= \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2m-1}{2}\right)} \frac{1}{(2\pi i)^m} \int_g \cdots \int \frac{(1 - \kappa_j r_j)^{\frac{1}{2}(n-2m-3)}}{\prod_{l=1}^{\frac{1}{2}(n-1)} (1 - \kappa_j \kappa_{jl})} \, d\kappa_1 \cdots d\kappa_m \\ &= \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-2m-1}{2}\right)} \sum_{g_k} \frac{\left| \begin{matrix} 1 & I \\ r & K_k \end{matrix} \right|^{\frac{1}{2}(n-2m-3)}}{\prod_{l \neq g_k} \left| \begin{matrix} 1 & I \\ \kappa_{jl} & K_k \end{matrix} \right|} , \end{aligned} \tag{12}$$

where $I = (1, 1, \cdots, 1)$, $r' = (r_1 , r_2 , \cdots , r_m)$, $\underline{\kappa}_{jl} = (\kappa_{jl} , \cdots , \kappa_{ml})$ and K_k is the matrix formed from a set g_k of the m matrices $\underline{\kappa}_{jl}$ arranged in order. The factors in the summation can most easily be determined if we put $\left| \begin{matrix} 1 & I \\ r & K_k \end{matrix} \right| \propto A(r_1 , \cdots , r_{m-1}) - r_m$ and sum over the region for which $r_m \leq A(r_1 , \cdots , r_{m-1})$. To demonstrate the manner in which formula (12) works, we shall consider $m = 2$. From formula (2) we can see that a limit to the possible values that r_2 can take is given by $r_2 = 2r_1^2 - 1$ i.e. by the curve $(\cos \theta, \cos 2\theta)$

[1] This is a sufficient condition for $x_{n+1} = x_1$.

in the (r_1, r_2) plane. It is not difficult to see that there are ${}^nC_2$ possible terms in (12) and that each of these terms is proportional to the $\frac{1}{2}(n - 2m - 3)$th power of the distance from a line in the (r_1, r_2) planes. These lines are the joins of the points $(\cos 2\pi i/n, \cos 4\pi i/n)$, $i = 1, \cdots, \frac{1}{2}(n-1)$ and the joins of such points on the curve $(\cos\theta, \cos 2\theta)$ give the outer limits of the possible values of r_1 and r_2. It can also be seen that these points correspond to the equations $\kappa_j\kappa_{jl} = 1$ (each of these equations determines a plane in 4-dimensional complex space), while the joins of these points correspond to the singularities defined by and terms arising from pairs of these equations. Furthermore, since the sum of residues in any plane is zero, the sum of contributions, taken with appropriate signs, arising from lines through any of these points is zero, i.e. the sum of all possible terms involving any particular κ_{jl} will disappear. This leads to several possible expansions for $h(r_1, \cdots, r_m)$.

If we consider the particular case $n = 9$, then each term in the expansion (12) is proportional to the distance from one of the lines joining $(\cos 2\pi i/9 \cos 4\pi i/9)$, $i = 1, 2, 3, 4$ These lines may be denoted by l_{ij}. Then the contribution from l_{ij} is given by

$$3\,\frac{\kappa_{1i}\kappa_{1j} - (\kappa_{1i} + \kappa_{1j})r_1 + \frac{1}{2}(r_2 + 1)}{(\kappa_{1j} - \kappa_{1i})(\kappa_{1i} - \kappa_{1k})(\kappa_{1i} - \kappa_{1i})(\kappa_{1j} - \kappa_{1k})(\kappa_{1j} - \kappa_{1i})},$$

where $j > i$ and $\kappa_{1\alpha} = \cos\dfrac{2\pi\alpha}{q}$.

The values of this expression are:

$$\begin{aligned}
l_{12}, &\quad -1.979 + 2.938\,r_1 - 1.563\,r_2,\\
l_{13}, &\quad 0.926 - 2.106\,r_1 + 3\,959\,r_2,\\
l_{14}, &\quad 1\,053 - 0.832\,r_1 - 2\,396\,r_2,\\
l_{23}, &\quad -5.012 - 3.959\,r_1 - 6.065\,r_2,\\
l_{24}, &\quad 3.033 + 6\,897\,r_1 + 4.502\,r_2,\\
l_{34}, &\quad -4.086 - 6.065\,r_1 - 2\,106\,r_2,
\end{aligned}$$

where, for example, the contribution from l_{12} acts in the region for which $1.563\,r_2 \leq -1\,979 + 2.938\,r_1$. Fig. 2 demonstrates the configuration for this case It is seen that the frequency surface is a tetrahedron. As particular examples of the identities mentioned above we have

$$\begin{aligned}
l_{12} + l_{13} + l_{14} &= 0,\\
-l_{12} + l_{23} + l_{24} &= 0,\\
-l_{13} - l_{23} + l_{34} &= 0.
\end{aligned}$$

For a general value of m, we shall find that the hyperplanes joining sets of m points $(\cos 2\pi i/n, \cos 4\pi i/n, \cdots, \cos 2\pi mi/n)$ will be singularities on the

frequency hypersurface. The hyperplanes passing through sets of m successive points will give the limits of possible values of $r_1, \cdots, r_m$. Furthermore, the sum of contributions (with appropriate signs) to the frequency function from the set of $\frac{1}{2}(n - 2m + 1)$ hyperplanes passing through any point will be zero.

8. Integral approximation for the distribution function. The expression (12) is, of course, difficult to use in practice and we require an approximation similar to that of Koopmans. For this we make use of the integral expression (10)

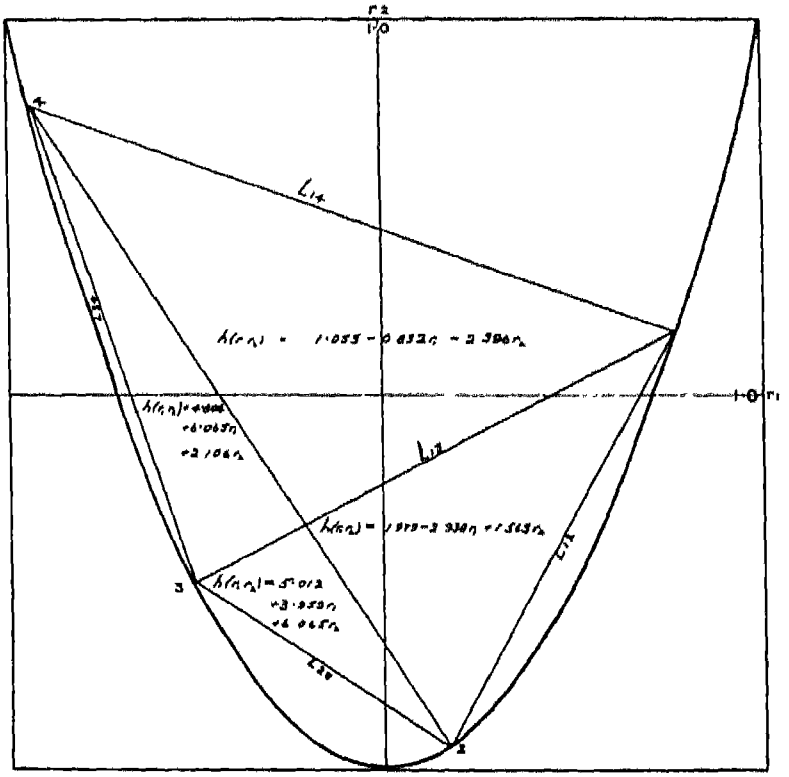

FIG. 2

for the joint distribution function of $\bar{r}_1, \cdots, \bar{r}_m$ and approximate to the factor $\left[\prod_{i=1}^{n}(1 - \kappa_j\kappa_{jl})\right]^{-\frac{1}{2}}$. This can be done without undue difficulty, but the resulting multiple integral does not appear to be capable of easy reduction. This is hardly surprising, since from the nature of the distribution of the r_i we should expect this approximation to involve R_m raised to a suitable power, and this conjecture is strengthened by the following considerations:

a) The distribution of $\bar{r}_1$ may be obtained by considering the two sets of observations $x_1, x_2, \cdots, x_{n-1}, x_n$ and $x_2, x_3, \cdots, x_n, x_1$ as unrelated, and using

the distribution of the ordinary correlation coefficient corresponding to $n + 3$ pairs of observations. (Dixon [6] Quenouille [7]). In the same manner, the m sets of observations $x_1, x_2, \cdots, x_{n-1}, x_n; x_2, x_3, \cdots, x_n; \cdots x_m, x_{m+1}, \cdots, x_{m-2}, x_{m-1}$, might be considered as unrelated and the joint distribution of their correlations, given by Garding (5), will involve R_m raised to a suitable power.

b) The outer limits for the joint distribution of $r_1, r_2, \cdots, r_m$ or $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_m$ for large n, will be provided by the equations $R_p = 0$, $(p = 1, \cdots, m)$. An investigation of the properties of the functions, $R_1, R_2, \cdots, R_m$ might therefore be expected to throw light upon the joint distribution of $r_1, r_2, \cdots, r_m$ or $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_m$.

c) R_p is a quadratic in r_p and may be put equal to $R_{p-2}(r'_p - r_p)(r_p - r''_p)$, where r'_p and r''_p are functions of $r_1, r_2, \cdots, r_{p-1}$ giving the limits of the values that r_p can take for any particular values of $r_1, \cdots, r_{p-1}$. Let $Q_p = R_p/R_{p-1}$, then Q_p is likewise a quadratic in r_p, taking all values between r'_p and r''_p and

$$\int_{r''_p}^{r'_p} Q_p^s \, dr_p = \frac{R_{p-2}^s}{R_{p-1}} \int_{r''_p}^{r'_p} (r'_p - r_p)^s (r_p - r''_p)^s \, dr_p$$

$$= \frac{B(s+1, \frac{1}{2})}{Q_{p-1}^s} \cdot \left(\frac{r'_p - r''_p}{2}\right)^{2s+1}.$$

But, by expanding R_p as a bordered determinant, it is not difficult to show that $r'_p - r''_p = 2Q_{p-1}$, so that

$$\int_{r''_p}^{r'_p} Q_p^s \, dr_p = \frac{\Gamma(s+1)}{\Gamma(s+\frac{3}{2})} \cdot \pi^{\frac{1}{2}} \cdot Q_{p-1}^{s+1}$$

In particular, if

$$(13) \qquad f(r_1 \cdots r_m) = \frac{\Gamma(\frac{1}{2}n+1)\cdots\Gamma(\frac{1}{2}n-m+2)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\cdots\Gamma(\frac{1}{2}n-m+\frac{3}{2})} \cdot \frac{1}{\pi^{m/2}} Q_m^{\frac{1}{2}(n-2m+1)},$$

and if we integrate with respect to $r_m, r_{m-1}, \cdots r_2$ in turn, we get

$$\int_{r''_2}^{r'_2} \cdots \int_{r''_m}^{r'_m} f(r_1 \cdots r_m) \, dr_m \cdots dr_2 = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\pi^{\frac{1}{2}}} (1-r^2)^{\frac{1}{2}(n-1)},$$

which is the approximate distribution of the first serial correlation coefficient, uncorrected for the mean, as given by Dixon [6].

The importance of this lies in the fact that the integral corresponding to that of Koopman's for the joint distribution is

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n-m)} \cdot \frac{2^{\frac{1}{2}n}}{\Pi^m} \int_s \cdots \int \frac{\left|\begin{matrix} 1 & \underline{I} \\ r & \mathbf{X} \end{matrix}\right|^{\frac{1}{2}n-m-1}}{|\mathbf{Y}|^{\frac{1}{2}n}} \prod_{i=1}^{m} \left[\sin \tfrac{1}{2}nx_i \left| \begin{matrix} 0 & \\ \frac{d}{dx_i} \psi(x_i) & \begin{matrix} I \\ X \end{matrix} \end{matrix} \right| \right] dx_1 \cdots dx_m$$

where $r' = [r_1, \cdots, r_m]$,

$$\mathbf{X} = \begin{bmatrix} \cos x_1 & \cos x_2 & \cdots & \cos x_m \\ \cos 2x_1 & \cos 2x_2 & \cdots & \cos 2x_m \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \cos mx_1 & \cos mx_2 & \cdots & \cos mx_m \end{bmatrix},$$

$$\mathbf{I} = [1, 1, \cdots 1],$$

$$\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos x_1 & \cos x_2 & \cdots & \cos x_m \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \cos (m-1)x_1 & \cos (m-1)x_2 & \cdots & \cos (m-1)x_m \end{bmatrix},$$

$$\kappa'(\theta) = [\cos\theta, \cos 2\theta, \cdots, \cos m\theta],$$

and S is the region given by $\begin{vmatrix} 1 & I \\ r & X \end{vmatrix} \geq 0$. This suggests, by analogy, that the joint distribution function is a polynomial in r_m of degree $2(\frac{1}{2}n - m - 1) + 3 = n - 2m + 1$ which vanishes only when $R_m = 0$. The equation satisfies these conditions, and in addition, it reduces to the known form when $m = 1$ and can be integrated to give this same form. Thus there is a strong suggestion that (13) gives an approximate distribution of $r_1, r_2, \cdots, r_m$, uncorrected for the mean.

An alternative form for the constant factor in (13) may be obtained if we note that

$$\frac{\Gamma(\frac{1}{2}n - m + 2)}{\Gamma(\frac{1}{2}n - m + \frac{5}{2})\pi^{\frac{1}{2}}} = \frac{1}{2^{n-2m+2}} \frac{\Gamma(n - 2m + 3)}{[\Gamma(\frac{1}{2}n - m + \frac{5}{2})]^2}.$$

d) Now r'_p and r''_p can be written in the forms $(S_{p-1} + R_{p-1})/R_{p-2}$ and $(S_{p-1} - R_{p-1})/R_{p-2}$, where

$$S_{p-1} = (-1)^{p-1} \begin{vmatrix} r_1 & r_2 & r_3 & \cdots & 0 \\ 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ \cdots\cdots & \cdots\cdots & & & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & r_1 \end{vmatrix}.$$

Thus

$$R_p = R_{p-2}\left(\frac{S_{p-1} + R_{p-1}}{R_{p-2}} - r_p\right)\left(r_p - \frac{S_{p-1} - R_{p-1}}{R_{p-2}}\right)$$

$$= \frac{R_{p-1}^2}{R_{p-2}}\left[1 - \left(\frac{r_p R_{p-2} - S_{p-1}}{R_{p-1}}\right)^2\right]$$

$$Q_p = Q_{p-1}(1 - r_{1p\cdot 1\,23\cdots}^2)$$

where

$$r_{1,p+1.23\cdots} = T_{p-1}/R_{p-1},$$

and

$$T_{p-1} = \begin{vmatrix} r_p & r_1 & r_2 & \cdots & r_{p-1} \\ r_{p-1} & 1 & r_1 & \cdots & r_{p-2} \\ r_{p-2} & r_1 & 1 & \cdots & r_{p-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_{p-2} & r_{p-3} & \cdots & 1 \end{vmatrix}.$$

Therefore, if we make a change of variable to $r_{1,p-1.23\ldots}, r_{1,p.23\ldots}, \cdots r_{13\,2}, r_1$, we find that the new variables which correspond exactly to partial correlation coefficients are, in fact, independently distributed as such, with 3 degrees of freedom more than in the case where the sets of variables are distinct observations.

While the above properties do not prove that the r_i or $\bar{r}_i$ may be tested using partial or multiple correlation coefficients, this conjecture has been verified elsewhere and it has been shown [8] that, with certain adjustments, a test can be derived which is applicable to fairly short series.

REFERENCES

[1] R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Math Stat.*, Vol. 13 (1942), pp. 1–13

[2] T. Koopmans, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13 (1943), pp. 14–33

[3] M. S. Bartlett, "On the theoretical specification of sampling properties of autocorrelated time series," *Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp 27–41.

[4] W. G. Madow, "Note on the distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308–310

[5] L. Garding, *Proceedings of Lund University Mathematical Seminars*, Vol. 5, pp 185–202

[6] W. J. Dixon, "Further contributions to the problem of serial correlation," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[7] M. H. Quenouille, "Some results in the testing of the serial correlation coefficient," *Biometrika*, Vol. 35 (1948), pp. 261–7.

[8] M. H. Quenouille, "Approximate tests of correlation in time series 1," *Roy Stat Soc Suppl.*, Vol. 11 (1949).

ON THE ESTIMATION OF THE NUMBER OF CLASSES IN A POPULATION[1]

By Leo A. Goodman

Princeton University

1. Summary. This paper deals with the following problem: Suppose a population of known size N is subdivided into an unknown number of mutually exclusive classes. It is assumed that the class in which an element is contained may be determined, but that the classes are not ordered. Let us draw a random sample of n elements without replacement from the population. The problem is to estimate the total number K of classes which subdivide the population on the basis of the sample results and our knowledge of the population size.

There is exactly one real valued statistic S which is an unbiased estimate of K when the sample size n is not less than the maximum number q of elements contained in any class. The restriction placed upon q is unimportant for many practical problems where either there is a reasonably low bound for q or those classes containing more than n elements are known. An unbiased estimate does not exist when there is no such knowledge.

Since the unbiased estimate can be very unreasonable, modifications of S are considered. The statistic

$$T' = \begin{cases} S' = N - \dfrac{N(n-1)}{n(n-1)}\, x_2, & \text{if } S' \geq \sum_{i=1}^{n} x_i, \\ \sum_{i=1}^{n} x_i, & \text{if } S' < \sum_{i=1}^{n} x_i, \end{cases}$$

where x_i is the number of classes containing i elements in the sample, is the most suitable estimate, in comparison with three other statistics, for a hypothetical population.

The case where each element in the population has an equal and independent chance of coming into the sample is used as a model for some sampling procedures and also as an approximation to the case of random sampling.

2. Introduction. The problem discussed may be described in terms of colored balls in an urn. How should we estimate the number of colors present in the urn on the basis of both the sample which gives the number of, say, white balls, red balls, etc., and our knowledge of the total number of balls in the urn:

The following practical cases illustrate some of the ways in which this problem presents itself:

(1) A company has received a large number of requests for a free sample of its product. It is known that the same people often send more than one request.

[1] Prepared in connection with research sponsored by the Office of Naval Research.

From a sample of the requests we wish to estimate how many different people have sent requests.[2]

(2) The Social Security Board possesses a large collection of Social Security cards. It is known that some people obtain different cards when they change jobs. From a sample of the cards it is desired to estimate how many different people have Social Security cards.[3]

(3) A person who sells durable commodities anticipates opening a store which is to be located at a highway intersection. He would like to know how many different automobiles pass through the intersection in a given time period. The total number of automobiles may be easily observed but some probably pass through more than once. This type of inquiry is also useful to advertising agencies which must decide the most efficient location for billboards.

(4) The State Unemployment Compensation Board possesses a large list of the people receiving unemployment benefits. It is desired to estimate the total number of families benefiting from the insurance program on the basis of a random sample of the people named on the list.

(5) The number of words in a book may be easily estimated and a sample can be taken. The problem of estimating the number of different words in a book is another analogue of the general problem.[4]

3. Results and derivations. In order to show that an unbiased estimate of the number of classes in a population exists when the sample size n is not less than the maximum number q of elements contained in any class, we need prove the following two statements:

LEMMA 1. *Suppose we have K classes of N similar elements with n_1 elements in class 1, n_2 elements in class 2, $\cdots$, n_K elements in class K. The class of an element is readily identifiable when the element is examined. Let*

$$q = \max(n_i).$$

Suppose a random sample is drawn without replacement. If x_i is the number of classes containing i elements in the sample, and K_j is the number of classes containing j elements in the population, then

$$E(x_i) = \sum_{j=1}^{q} \Pr(i \mid j, N, n) K_j,$$

where $Pr(i \mid j, N, n)$ shall henceforth be an abbreviation of

$$\frac{C_i^j C_{n-i}^{N-j}}{C_n^N}.$$

[2] Submitted by Charles Callard to Question and Answers, *The American Statistician*, Vol. 3, No. 1, p. 23.

[3] Mentioned to the author by Dr. J. Stevens Stock of Opinion Research Corporation

[4] Mentioned in letter to the author from Frederick Mosteller of Harvard University.

PROOF. Let y_s be the number of elements appearing in the sample from the s-th class. The statement is proved by considering $E(x_i) = \sum_{s=1}^{K} E(\delta_{iy_s})$, where

$$\delta_{iy_s} = \begin{cases} 1, & \text{if } y_s = i, \\ 0, & \text{if } y_s \neq i. \end{cases}$$

LEMMA 2. *Let*

$$a^{(t)} = \begin{cases} a(a-1)(a-2)\cdots(a-t+1), & \text{for } t > 0, \\ 1, & \text{for } t = 0. \end{cases}$$

If

$$a_i = 1 - (-1)^i \frac{[N-n+i-1]^{(i)}}{n^{(i)}},$$

then

$$\sum_{i=1}^{j} A_i \Pr(i \mid j, N, n) = 1.^5$$

This result follows directly from the fact that

$$\sum_{i=0}^{j} (-1)^i C_i^j [N-n+i-1]^{(j-1)} = 0, \text{ for } j \geq 1.$$

The following theorem may be proved directly by the preceding lemmas:

THEOREM 1. *Suppose a sample of n elements is drawn without replacement from a population of size N which is subdivided into K classes. Let*

$$A_i = 1 - (-1)^i \frac{[N-n+i-1]^{(i)}}{n^{(i)}}.$$

If there are x_i classes containing i elements in the sample, then

$$E\left(\sum_{i=1}^{n} A_i x_i\right) = K,$$

provided that n is not less than the maximum number q of elements contained in any class in the population.

THEOREM 2. *There is at most one real valued statistic which is an unbiased estimate of the number of classes in a population.*[6]

PROOF. Let us order the points of the sample space in the following manner: Letting x_i be the number of classes containing i elements in the sample, order the sample points by increasing values of x_n; for equal values of x_n, order the points by increasing values of x_{n-1}; for equal values of x_{n-1}, order the points

[5] The author is indebted to Professor Frederick F. Stephan of Princeton University for a statement leading to a simplification of the original result.

[6] This statement was mentioned to the author by M. P. Peisakoff of Princeton University.

by increasing values of x_{n-2} ; $\cdots$; for equal values of x_3, order the points by increasing x_2. Let

$$x_1 = n - \sum_{j=2}^{n} jx_j .$$

To prove the theorem, we must show that to each 0_i there corresponds a unique value $S(i)$, which must be the value of our estimate when 0_i is observed, in order that the statistic be unbiased. To each

$$0_i = [x_1(i), x_2(i), x_3(i), \cdots, x_n(i)],$$

let us associate the population

$$P_i = \left[N - \sum_{j=2}^{n} jx_j(i), x_2(i), x_3(i) \cdots, x_n(i) \right].$$

If P_1 is the underlying population, then 0_i for all $i > 1$ will occur with a probability of zero. Since there are N classes in P_1, the value of the statistic must be $S(1) = N$ whenever 0_1 is observed in order that the estimate be unbiased. The theorem may now be proved by induction.

Since all the P_i used in the proof of Theorem 2 satisfied the condition that the maximum number q of elements contained in any class be not more than the sample size n, the statistic S is the only real valued statistic which is an unbiased estimate when $q \leq n$.

When the restriction that $q \leq n$ is removed, it is useless to search for an unbiased estimate since we have

THEOREM 3. *There does not exist an unbiased estimate of the number of classes subdividing a population when it is not known whether the maximum number q of elements contained in any class is not more than the sample size n.*

By the preceding theorems it is clear that if an unbiased estimate exists it must equal S. However, S is generally not unbiased when $n < q$.

THEOREM 4. *Suppose the statistics $S_1, S_2, \cdots, S_n$ are the solutions of the system of linear equations*

$$x_i = \sum_{j=1}^{n} \Pr(i \mid j, N, n) S_j , \text{ for } i = 1, 2, \cdots, n,$$

where x_i is the number of classes containing i elements in a sample of size n from a population of N elements. If K_j is the number of classes containing j elements in the population, then $E(S_j) = K_j$, for $j = 1, 2, \cdots, n$ when n is not less than the maximum number q of elements contained in any class.

PROOF. We observe that the statement is certainly true for $j = q + 1, q + 2, \cdots, n$, since

$$E(S_j) = K_j = 0, \quad \text{for} \quad j = q + 1, q + 2, \cdots, n.$$

The statement is also true for $j = q$, since

$$E(S_q) = E(x_q) \frac{N^{(q)}}{n^{(q)}} = K_q .$$

To prove that $E(S_j) = K_j$, for any $j < q$, we assume it to be true for all $i > j$, whereupon its truth for j follows.

By Theorem 2, and 3, it is clear that $\sum_{j=1}^{n} S_j = S$. Since

$$\sum_{j=1}^{q} jK_j = N,$$

it seems reasonable to ask whether the values of the estimates $S_1, S_2, \cdots, S_n$ are in agreement with the known value of the size of the population. The unbiased estimate of K can be shown to be internally consistent by

THEOREM 5. *Suppose a sample of size n is drawn without replacement from a population of N elements which is divided into classes. If x_i is the number of classes containing i elements in the sample, and if the linear equations*

$$x_i = \sum_{j=1}^{n} \Pr(i \mid j, N, n) S_j,$$

are solved simultaneously for S_j, then

$$\sum_{j=1}^{n} jS_j = N.$$

The theorem follows readily from the fact that

$$\sum_{i=1}^{j} i \Pr(i \mid j, N, n) = n \frac{j}{N} \text{ and } \sum_{i=1}^{n} ix_i = n.$$

The variance of S may now be calculated by means of the formula

$$\sigma_S^2 = \sum_{i,j=1}^{n} A_i A_j u_{ij} = \sum_{i,j=1}^{n} A_i A_j \left\{ \sum_{s,t=1}^{q} m_{st}(i,j) K_s K_t + \sum_{s=1}^{q} [m_s(i,j) - m_{ss}(i,j)] K_s \right\},$$

where u_{ij} is the covariance between x_i and x_j, $m_{st}(i,j)$ is the covariance between δ_{iv_r} and δ_{jv_h} when $r \neq h$, $n_r = s$ and $n_h = t$, and $m_s(i,j)$ is the covariance between δ_{iv_r} and δ_{jv_r} when $n_r = s$.

Since the statistic S can be very unreasonable, we consider other possible estimates of K. The statistic

$$S' = N - \frac{N^{(2)}}{n^{(2)}} x_2$$

may be shown to be a modification of S which replaces the number x_i of classes containing $i > 2$ elements in the sample by an additional ix_i classes, each containing only one element. Since the values of K_i for $i > 2$ are relatively small in the practical problems of Section 2, S' might be used as an estimate.

Another statistic which may be used to estimate K is

$$S'' = \frac{N}{n} \sum_{i=1}^{n} x_i.$$

This statistic may be shown to overestimate K whenever $q \neq 1$. The estimate

$$S'' = \sum_{i=1}^{n} x_i$$

underestimates K when $n \leq N - m$ where m is the least number of elements contained in any class.

4. Binomial sampling. Let us suppose that each element from a population of N elements has an equal and independent chance $p = 1/r$ of entering the sample s. In this case, the size of the sample obtained is a random variable η which is binomially distributed with mean Np. If a large random sample of n elements is drawn without replacement from a large population of size N, then the results when interpreted in terms of binomial samples where $p = 1/r = n/N$ are a good approximation to the results obtained by the usual model. Binomial sampling may be considered a model of the case where one attempts to obtain the sampling ratio $p = 1/r$ by drawing simultaneously an uncounted sample of elements which is estimated as being of the appropriate size.

In the case of binomial sampling, the statistic

$$B = \sum_{i=1}^{N} B_i x_i, \quad \text{where} \quad B_i = 1 - (1 - r)^i$$

may be shown to be an unbiased estimate of the number of classes in a population from which binomial samples are drawn.

Let us now consider the statistic which corresponds to S' for the case of binomial sampling; i.e.,

$$B' = N - r^2 x_2 .$$

It may be shown that

$$E(B') = K_1 + K_2 + \sum_{j=3}^{q} [j - C_2^j (1 - p)^{j-2}] K_j .$$

Hence, the statistic B' will underestimate K whenever

$$p < 1 - \left(\frac{2}{j}\right)^{1/j-2}, \quad \text{for} \quad j = 3, 4, \cdots, q.$$

Since

$$1 - \left(\frac{2}{j}\right)^{1/j-2}$$

is a decreasing function of j for $j > 2$, when $p > \frac{1}{3}$, B' overestimates, and when

$$p < 1 - \left(\frac{2}{p}\right)^{1/q-2},$$

B' underestimates the value of K. When p is such that

$$1 - \left(\frac{2}{q}\right)^{1/q-2} \leq p \leq \tfrac{1}{2},$$

the expected value of B' is brought closer to K by underweighting some K_j and overweighting others.

5. A hypothetical population.[7] Suppose we draw a random sample of 1000 elements without replacement from a population of 10,000 elements where

$$K_1 = 9225, \qquad K_2 = 336, \qquad K_3 = 33, \qquad K_4 = 1.$$

Hence, $K = 9595$. By means of Table 1, let us now compare on the basis of binomial sampling the estimates which have been presented in the preceding sections. Since N and n are large, these results are a good approximation to the case of random sampling without replacement.

TABLE 1

Estimate	*Expected value*	*Bias*	$\sqrt{\textit{Mean Square Error}}$
S	9595	0	347
S'	9570	−25	207
S''	9959	364	490
S'''	996	−8599	8600

It is clear that the best estimates of the number of classes in this particular population are S or S', since S has the least bias, $E(S) - K$, and S' has the least mean square error, $E(S' - K)^2$. One might argue that both S and S' are the statistics which are capable of giving nonsensical estimates. However, we may decide to modify S or S' in order to always get reasonable estimates by using the statistics

$$T = \begin{cases} S, & \text{if } N \geq S \geq \sum_{i=1}^{n} x_i, \\ N, & \text{if } S > N, \\ \sum_{i=1}^{n} x_i, & \text{if } S < \sum_{i=1}^{n} x_i \end{cases}$$

$$T' = \begin{cases} S', & \text{if } S' \geq \sum_{i=1}^{n} x_i, \\ \sum_{i=1}^{n} x_i, & \text{if } S' < \sum_{i=1}^{n} x_i. \end{cases}$$

[7] Other examples have been investigated by Frederick Mosteller in Questions and Answers, *The American Statistician*, Vol. 3, No. 3, p. 12.

Although these modified statistics T and T' are not unbiased, they have the desirable property that

$$MSE(T) \leq MSE(S), \quad \text{and} \quad MSE(T') \leq MSE(S').$$

Since this hypothetical population is a plausible one for the practical problems of Section 2, the modified statistics T or T' seem, therefore, to be "best" for estimating the number of classes for these problems, where the "best" statistic is defined as the one which never gives unreasonable estimates and has the least mean square error.

The author wishes to express his appreciation to Professor John W. Tukey whose suggestions were very helpful.

CONCERNING COMPOUND RANDOMIZATION IN THE BINARY SYSTEM

By John E. Walsh

The Rand Corporation

1. Summary. Let us consider a set of approximately random binary digits obtained by some experimental process. This paper outlines a method of compounding the digits of this set to obtain a smaller set of binary digits which is much more nearly random. The method presented has the property that the number of digits in the compounded set is a reasonably large fraction (say of the magnitude $\frac{1}{3}$ or $\frac{1}{4}$) of the original number of digits.

If a set of very nearly random decimal digits is required, this can be obtained by first finding a set of very nearly random binary digits and then converting these digits to decimal digits.

The concept of "maximum bias" is introduced to measure the degree of randomness of a set of digits. A small maximum bias shows that the set is very nearly random.

The question of when a table of approximately random digits can be considered suitable for use as a random digit table is investigated. It is found that a table will be satisfactory for the usual types of situations to which a random digit table is applied if the reciprocal of the number of digits in the table is noticeably greater than the maximum bias of the table.

2. Introduction and discussion. With the development of the theory of games and the more widespread use of experimental methods for determining approximate distributions for statistics whose probability laws are difficult to obtain analytically, a demand for large sets of random digits has arisen. The problem of obtaining a set of digits which can be considered sufficiently random for the situations to which it would be applied, however, is not an easy one. One approach to this problem consists in obtaining a set of digits by some procedure and then applying tests to this set of digits to determine whether it can be considered satisfactory. Although appropriate choice of the tests may result in acceptance of sets of digits which are suitable for certain special types of situations, this approach is of a negative character and does not prove that a given set of digits is sufficiently random; it merely indicates that this may be the case. What is needed is a constructive approach to the problem, i.e., a method of constructing a set of random digits which can be proved sufficiently random for most applications if certain intuitively acceptable conditions are satisfied. A step in this direction has already been taken by H. Burke Horton in [1] and by H. Burke Horton and R. Tynes Smith III in [2]. This paper presents what is hoped will be another step in this direction.

In this paper, considerations will be limited to the case of binary digits. The reasons for this are twofold:

(a). The method used for compounding the digits yields a sharp upper bound for the maximum bias of the compounded set (i.e., a bound that the maximum bias could actually attain) only for the case of binary digits.

(b). Many of the experimental procedures for obtaining approximately random digits consist in first producing binary digits and then converting to another number base. Thus binary digits are produced directly. Hence, to use the results of this paper, the only modification required in these procedures would be to compound the binary digits before they are converted.

Now let us consider some definitions: A set of random variables each of which can assume only the values 0 and 1 will be referred to as a set of binary digits. For convenience, each of the random variables making up a set of binary digits will be called a binary digit; this is not to be confused with the value obtained for the random variable. The absolute value of the deviation from $\frac{1}{2}$ of the conditional probability that a specified binary digit has the value 0 (or 1) is called the *bias* of that digit for the given conditions on the remaining digits of the set. The maximum bias of a binary digit is defined to be the maximum of the biases of that digit with respect to all possible conditions on the remaining digits of the set. The *maximum bias of the set* is the greatest of the maximum biases of the digits of the set. A set of binary digits is said to be *random* if its maximum bias is zero.

The method used to prove that a set of compounded digits has a sufficiently small maximum bias is somewhat similar to the situation encountered in mathematics where one begins with certain axioms and then draws conclusions. If the axioms are correct, the conclusions are necessarily valid. The first step in the compounding procedure consists in obtaining a set of binary digits by some experimental process (perhaps from a random digit machine which is based on some physical principle). The experimental process is so chosen that there is no doubt that the set of binary digits produced satisfies the two conditions:

(i). The maximum bias of the set is less than or equal to some specified value $\alpha(<\frac{1}{2})$.

(ii). The digits of the set can be arranged in a specified array which has the property that the rows of the array are statistically independent.

On the basis of these two assumptions (which play the same role as the axioms mentioned above), it can be proved that the maximum bias of the resulting compounded set of binary digits never exceeds a specified value which depends on α. Moreover, the upper bound for the maximum bias of the constructed set of binary digits can be made extremely small even for large values of α.

If the experimental process is suitably chosen, conditions (i) and (ii) can be satisfied beyond any doubt. For example, let us consider 1000 people located in different parts of the world and not in contact with each other. Let each person flip an ordinary coin high in the air so that it will land on a flat hard surface, record the result (say 0 for a tail and 1 for a head), and then repeat this procedure until 5000 binary digits are obtained. If α is set equal to 3/10, condition (i) is

obviously satisfied for the resulting set of 5,000,000 binary digits. Condition (ii) evidently holds if the array is taken to consist of 1000 rows where each row contains 5000 binary digits obtained from one person

The ideal choice for α would be the actual maximum bias of the set of binary digits obtained from the experimental process. Then the compounding procedure for obtaining a set of digits with a specified upper bound for the maximum bias would be simplified; also the number of digits in the compounded set would be a larger fraction of the original number of digits. Invariably, however, the properties of the experimental process are not known with sufficient accuracy for obtaining anything but a safe upper bound on the maximum bias of the set of digits produced. This situation is analogous to that of estimating the length of a stick which a very rough measurement has shown to be about 10″ long. Although one might be very hesitant to believe that the length of the stick lies between 9.9″ and 10.1″, the contention that the length lies between 5″ and 15″ can be accepted with virtual certainty and any logical conclusions based on this contention can also be accepted with virtual certainty.

Given the number of binary digits in a set and the maximum bias of the set, is it possible to determine whether the set is suitable for use as a set of random binary digits? An important consideration in answering this question is the use that is to be made of the set of digits. This must always be taken into account before the suitability of the set can be decided. For example, if no more than 1/1000 of the digits of the set are to be used for any particular situation, the set might be satisfactory for the types of cases to which it would be applied; on the other hand, the set might not be suitable for cases of these types if all the digits of the set are used for each situation. This example calls attention to an important point, namely that the suitability of a set of binary digits depends on the number of digits in the set. Let a set have a fixed non-zero maximum bias ρ. If the set contains a sufficiently large number N of digits, relations and expressions involving the digits of the set can be found whose probabilities, moments, etc., can differ greatly from the values which would be obtained if the relations were based on the same number of truly random binary digits. As a specific example consider the relation

All the digits of the set have the value zero.

If the reciprocal of the number of digits in the set is of the same order of magnitude or smaller than the maximum bias of the set, the ratio of the probability of this expression to its hypothetical value can differ noticeably from unity. Thus, at least in certain special cases, a necessary condition for the suitability of a set of binary digits is that $1/N > > \rho$. This condition, however, is also sufficient for most situations to which a set of random digits would be applied. The approximate sufficiency of the condition is a direct consequence of the fact that any set of N binary digits can be considered as a sample value from an N-dimensional population consisting of 2^N discrete points. The $1/N > > \rho$ restriction implies that the probability concentrated at each of the 2^N points is

very nearly equal to the hypothetical value of $(\frac{1}{2})^N$ for all possible conditions on the remaining digits of the set.

The $1/N > > \rho$ condition is very satisfactory from the viewpoint of probabilities. The probability of any relation based on a subset of the digits of the set (possibly conditioned on other digits from the table) can be interpreted as the sum of the probabilities of those points included in a certain region (defined by the relation) of the N-dimensional probability space of the set of digits. By expanding $(\frac{1}{2} \pm \rho)^N$ it can be shown that the ratio of the probability of any relation based on one or more digits from the set to the corresponding value for a truly random set of digits will be very nearly equal to unity if $1/N > > \rho$

It is evident that the higher order moments of an expression based on one or more digits of the set can differ noticeably from its hypothetical value even if $1/N > > \rho$; any deviation from the ideal situation, no matter how small, can become important for high order moments. For the first few moments, however, deviations from the hypothetical values are not appreciable since these moments are based on the probabilities at the 2^N points in the N-dimensional probability space and these probabilities are very nearly equal to the hypothetical value of $(\frac{1}{2})^N$ in all cases.

The above discussion shows that the values of N and ρ are sufficient to determine whether a set of binary digits is suitable for use as random binary digits for a wide variety of situations. Analogous considerations apply for digits to any number base.

A magnitude definition of the relation $1/N > > \rho$ is difficult to specify. If ρ is the upper bound for the maximum bias of a set of digits obtained by the compounding procedure outlined in this paper, however, it seems that a reasonable condition would be that $1/N \geq 50\ \rho$. This condition implies that the probability of any relation based on digits of the set can not differ from its hypothetical value by more than approximately 4%. In most practical applications the value obtained for ρ would be noticeably greater than the true value of the maximum bias of the compounded set.

Since the maximum number of digits which can be taken from a table is the total number of digits in the table, the above considerations suggest that a random digit table should be constructed so that the reciprocal of the number of digits in the table is noticeably greater than the maximum bias of the table. Any table having this property would be satisfactory for most situations to which it would be applied.

Now let us consider two different compounding methods which produce sets of binary digits with the same upper bound for the maximum bias. If the computational difficulties of applying the two methods are of comparable magnitudes, it seems reasonable to prefer the method which yields the larger set of digits. For example, if the number of digits in the set obtained by the first method is only 1/8 of the original number of digits while the number in the set obtained by the second method is 1/3 of the original number, the second method would seem preferable even if it required as much as 100% more computation.

The compounding method presented in this paper has the property that the number of digits in the compounded set can be held to a reasonably large fraction of the original number of digits at the same time that the upper bound for the maximum bias is made extremely small. The method presented by Horton in [1] does not have this property. For example, let $\alpha = 1/10$. Applying Horton's method, when the compounded set consists of 1/8 of the original number of digits the upper bound for the maximum bias is 12.8×10^{-7}. The example presented in section 3, however, shows that a compounded set whose number of digits equals 1/3 of the original number and which has an upper limit of 11.7×10^{-7} for the maximum bias can be obtained using the method presented in the next section.

Although the compounding method outlined in section 3 is presented as a series of steps, the value of a digit of the compounded set can be written as a linear function (mod 2) of digits of the original set. This was not done in what follows because of the complicated nature of the general form of such expressions. In any particular case, however, these expressions can be written without much trouble and the compounded digits computed from the original digits in a single step.

3. Outline of compounding method and statement of theorems. This section contains a description of the compounding method mentioned in the preceding two sections as well as statements of the basic theorems concerning this compounding method. Proofs of the results stated in this section are given in section 4.

Let us consider the array of mn binary digits

$$\begin{matrix} x_{11}, & x_{12}, & \cdots, & x_{1n} \\ x_{21}, & x_{22}, & \cdots, & x_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{m1}, & x_{m2}, & \cdots, & x_{mn} \end{matrix} \tag{1}$$

which satisfies conditions (i) and (ii); i.e., the maximum bias of the set (1) is less than or equal to α while a digit x_{uv} is independent of a digit x_{rs} if $r \neq u$ (if $r = u$, however, x_{uv} is not necessarily independent of x_{rs}).

Let a new set of $(m-1)n$ binary digits

$$y_{ij}, \qquad (i = 1, \cdots, m-1; j = 1, \cdots, n) \tag{2}$$

be formed as follows:

$$y_{ij} = x_{mj} + x_{ij} \pmod 2, \qquad (i = 1, \cdots, m-1; j = 1, \cdots, n).$$

Then the biases of the y_{ij} have the properties

THEOREM 1. *Let U be a specified set of $t-1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$, $(1 \leq t \leq m-1)$, while V is a specified set of zero or more of the y_{pq}'s*

with $q \neq j$. Also let θ consist of the set of integers such that $p \in \theta$ if $y_{pj} \in U$. Then, if γ_u = maximum bias for the set $x_{u1}, \cdots, x_{un}$, $(u = 1, \cdots, n)$,

$$|Pr(y_{ij} = 0 \mid U, V) - \tfrac{1}{2}| \leq \gamma_i [1 - \prod_{k \in \theta} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)] / [1 + \prod_{k \in \theta} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)]$$

for all possible selections of U, V and of the values for the digits of these sets.

COROLLARY 1. *If exactly $t - 1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$, have known values, the maximum bias of the binary digit y_{ij} is less than or equal to*

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^t/(\tfrac{1}{2} + \alpha)^t]/[1 + (\tfrac{1}{2} - \alpha)^t/(\tfrac{1}{2} + \alpha)^t].$$

COROLLARY 2. *The maximum bias of the set (2) is less than or equal to*

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^{m-1}/(\tfrac{1}{2} + \alpha)^{m-1}]/[1 + (\tfrac{1}{2} - \alpha)^{m-1}/(\tfrac{1}{2} + \alpha)^{m-1}].$$

The basic operation in the method of compounding binary digits is outlined in the procedure given for obtaining the y_{ij} from the x_{uv}. Let $m = (1 + t_1) \cdots (1 + t_K)$. Then a set of $t_1 \cdots t_K n$ binary digits can be obtained from the original set of mn digits x_{uv} by continually applying this basic procedure. The first step consists in dividing the rows of (1) into $(1 + t_2) \cdots (1 + t_K)$ sets each consisting of $(1 + t_1)$ rows in some specified fashion. Each of these sets is an array of $(1 + t_1) \times n$ binary digits for which the rows are independent. Apply the method used to obtain the y_{ij} from the x_{uv} to each $(1 + t_1) \times n$ array separately. Then each array yields a set of $t_1 n$ binary digits and there are $(1 + t_2) \cdots (1 + t_K)$ such sets. In each set arrange the $t_1 n$ digits into a single row in some specified manner. This furnishes a new array of $[(1 + t_2) \cdots (1 + t_K)] \times [t_1 n]$ binary digits for which the rows are independent. Repeat this procedure with respect to t_2 thus obtaining a new array of $[(1 + t_3) \cdots (1 + t_K)] \times [t_1 t_2 n]$ binary digits for which the rows are independent; etc., until a $(1 + t_K) \times (t_1 \cdots t_{K-1} n)$ binary digit array for which the rows are independent is obtained. Then form a set of binary digits Y_{gh}, $(g = 1, \cdots, t_K; h = 1, \cdots, t_1 \cdots t_{K-1} n)$, from this array in exactly the same manner that the y_{ij} were obtained from the x_{uv}. Then the biases of the Y_{gh} have the properties

THEOREM 2. *Let $\beta_0, \beta_1, \cdots, \beta_K$ be defined by $\beta_0 = \alpha$ and*

$$\beta_w = \beta_{w-1}[1 - (\tfrac{1}{2} - \beta_{w-1})^{t_w}/(\tfrac{1}{2} + \beta_{w-1})^{t_w}]/[1 + (\tfrac{1}{2} - \beta_{w-1})^{t_w}/(\tfrac{1}{2} + \beta_{w-1})^{t_w}], \quad (w = 1, \cdots, K)$$

Then, if exactly $t - 1$ of $Y_{1h}, \cdots, Y_{(g-1)h}, Y_{(g+1)h}, \cdots, Y_{t_K h}$ have known values, $(1 \leq t \leq t_K)$, the maximum bias of the digit Y_{gh} is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t].$$

In particular, the maximum bias of the entire set of Y_{gh} is less than or equal to β_K. Also

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t] \leq 2^{2^K - 1} \cdot t \cdot t_{K-1}^2 \cdot t_{K-2}^4 \cdots t_2^{2^{K-2}} \cdot t_1^{2^{K-1}} \cdot \alpha^{2^K}. \tag{3}$$

The inequality (3) is frequently useful from a computational viewpoint. Although the right hand side of (3) is usually noticeably greater than the left hand side, in many cases this rough upper bound is itself small enough to show that the upper bound for the maximum bias is of the desired order of magnitude.

If the set of compounded digits is to be used for a random binary digit table, Theorem 2 shows that advantage can be taken of the position of the digits in the table. Let $M = t_1 \cdots t_{K-1}n$ and enter the values of the Y_{gh}, $(g = 1, \cdots, t_K;$ $h = 1, \cdots, M)$, into the table in the order

$$Y_{11}, Y_{12}, \cdots, Y_{1M}, Y_{21}, \cdots, Y_{2M}, Y_{31}, \cdots, Y_{t_K1}, \cdots, Y_{t_KM}.$$

Then, if a set of digits is taken from this table in consecutive order (Y_{11} follows Y_{t_KM}), the upper bound for the maximum bias of this set is dependent on the number L of digits in the set. From Theorem 2, the maximum bias of a set of L digits taken in consecutive order from a table formed in this manner is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]$$

for values of L such that $(t - 1)M < L \leq tM$, where $1 \leq t \leq t_K$. Thus, if a small set of digits is taken from this table in consecutive order, the upper bound for the maximum bias of this set will usually be noticeably smaller than the upper bound for the maximum bias of the table. Since many uses of a random digit table require only a small fraction of the total number of entries in this table, this property would seem to be desirable. It should be emphasized, however, that the maximum bias of a set taken from this table is always less than or equal to β_K irrespective of the positions that the digits of the sets occupy in the table. Thus nothing is lost by constructing the table in this manner but something can be gained for small sets if the digits are taken from the table in consecutive order.

Now let us consider situations in which it is required that the number of digits in the compounded set is at least a specified fraction, say $1/C$, of the original number mn of binary digits. This requires that K and $t_1, \cdots, t_K$ be chosen so that

$$t_1 \cdots t_K/(1 + t_1) \cdots (1 + t_K) \geq 1/C.$$

Also, for given values of K and C, it seems preferable to choose $t_1, \cdots, t_K$ so that the value of β_K is at least approximately minimized. Examination of the results of Theorem 2 indicates that a reasonable method of determining the values of $t_1, \cdots, t_K$ with this in mind consists in first choosing t_1 as small as possible, then (given the value of t_1 equal to its minimum value) choosing t_2 as small as possible, etc. This method is also recommended by the fact that the resulting values of $t_1, \cdots, t_K$ are readily determined. The explicit procedure for finding $t_1, \cdots, t_K$ is given by

THEOREM 3. *Let the values of the integer K and the constant C (> 1) be given and consider the integers $t_1, \cdots, t_K$ subject to the condition*

$$t_1 \cdots t_K/(1 + t_1) \cdots (1 + t_K) \geq 1/C.$$

The minimum value of t_1 is the smallest integer satisfying

$$t_1 > 1/(C - 1).$$

In general, $2 \leq w \leq K - 1$, having already determined $t_1, \cdots, t_{w-1}$ as their minimum values, the value of t_w is the smallest integer satisfying

$$t_w > 1/[Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1].$$

Finally, given $t_1, \cdots, t_{K-1}$ as their minimum values, the minimum value of t_K is the smallest integer satisfying

$$t_K \geq 1/[Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1].$$

Now consider the general situation encountered in the application of the compounding process outlined above. Here the values of α, C are given and it is required to choose K and $t_1, \cdots, t_K$ so that the upper bound for the maximum bias of the compounded set of $t_1 \cdots t_K n$ binary digits Y_{gh} is less than or equal to a specified value b. The following procedure furnishes a method of solving this problem:

Let $K = 1$, obtain t_1 according to Theorem 3, and then compute β_1. If $\beta_1 \leq b$, a solution has been obtained. If $\beta_1 > b$, let $K = 2$ and repeat the procedure to obtain β_2. If $\beta_2 \leq b$, the values of t_1, t_2 and $K = 2$ are a solution. If $\beta_2 > b$, repeat the procedure for $K = 3$; etc. In practical situations, the value of K is usually bounded (e.g., by independence properties of the original set of digits). If β_K is still greater than b for the maximum permissible value of K, no solution is obtained. This means that either b must be increased or $1/C$ decreased or both if a solution is to be found. In many cases, a large amount of computation can be avoided by using the inequality (3). For marginal situations, however, a solution may be missed by using (3) instead of computing β_K.

Example of method. The following table represents an example of application of the above method:

$\alpha = 1/10$	$1/C = 1/3$	$b = 2 \times 10^{-6}$
$K = 1$, $t_1 = 1$		$\beta_1 = 2 \times 10^{-2}$
$K = 2$, $t_1 = 1$, $t_2 = 2$		$\beta_2 \leq 1.6 \times 10^{-3}$
$K = 3$, $t_1 = 1$, $t_2 = 3$, $t_3 = 9$		$\beta_3 \leq 1.04 \times 10^{-4}$
$K = 4$, $t_1 = 1$, $t_2 = 3$, $t_3 = 10$, $t_4 = 44$		$\beta_4 \leq 1.17 \times 10^{-6}$.

Thus $K = 4$, $t_1 = 1$, $t_2 = 3$, $t_3 = 10$, $t_4 = 44$ is a solution.

4. Derivations. The purpose of this section is to furnish proofs of the results stated in the preceding sections.

4.1 *Proof of Theorem* 1. Let us consider the conditional probability that an arbitrary but fixed y_{ij} has a specified value when the values of a fixed subset of zero or more of the remaining y's are known. For convenience, assume that y_{i1} is the binary digit considered and that the values of $y_{i2}, y_{i3}, \cdots, y_{it}$ (where t is a fixed integer such that $1 \leq t \leq m - 1$) and a set S are given while the

values of the remaining y's are unknown. Here S represents an arbitrary but fixed set of zero or more of the y_{ij}'s for which $j \geq 2$ while $t = 1$ has the interpretation that none of the y_{i1}, $(i \geq 2)$, are given. Let

$$Pr(x_{m1} = 0 \mid S) = \tfrac{1}{2} + \alpha_{t+1} \quad \text{and} \quad Pr(x_{k1} = b_k \mid S) = \tfrac{1}{2} + \alpha_k, \qquad (k = 1, \cdots, t).$$

Then, using the independence conditions satisfied by the x's,

$$\begin{aligned} Pr(y_{11} = b_1 \mid y_{21} = b_2, \cdots, y_{t1} = b_t; S) &= \left[\prod_{k=1}^{t+1}(\tfrac{1}{2} + \alpha_k) + \prod_{k=1}^{t+1}(\tfrac{1}{2} - \alpha_k)\right] \Big/ \left[\prod_{k=2}^{t+1}(\tfrac{1}{2} + \alpha_k) + \prod_{k=2}^{t+1}(\tfrac{1}{2} - \alpha_k)\right] \\ &= \tfrac{1}{2} + \alpha_1\left[\prod_{k=2}^{t+1}(\tfrac{1}{2} + \alpha_k) - \prod_{k=2}^{t+1}(\tfrac{1}{2} - \alpha_k)\right] \Big/ \left[\prod_{k=2}^{t+1}(\tfrac{1}{2} + \alpha_k) + \prod_{k=2}^{t+1}(\tfrac{1}{2} - \alpha_k)\right] \\ &= \tfrac{1}{2} + \alpha_1 \delta. \end{aligned}$$

Now $|\delta| = (1 - P)/(1 + P)$ if $0 \leq P \leq 1$ and equals $(P - 1)/(1 + P)$ if $P > 1$, where $P = \prod_{k=2}^{t+1}(\tfrac{1}{2} - \alpha_k)/(\tfrac{1}{2} + \alpha_k)$. Let γ_u be the maximum bias for the set of binary digits $x_{u1}, \cdots, x_{un}$, $(u = 1, \cdots, m)$. Then it is easily seen that

$$\max_P |\delta| \leq \left[1 - \prod_{k=2}^{t+1}(\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \Big/ \left[1 + \prod_{k=2}^{t+1}(\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right].$$

Thus

$$|Pr(y_{11} = b_1 \mid y_{21} = b_2, \cdots, y_{t1} = b_t; S) - \tfrac{1}{2}| \leq \gamma_1\left[1 - \prod_{k=2}^{t+1}(\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \Big/ \left[1 + \prod_{k=2}^{t+1}(\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right]$$

for all possible selections of $b_1, \cdots, b_t$ and all possible selections of S and the values for the digits of S. It is to be observed that this inequality is valid for $t = 1$.

Evidently this result can be modified to apply to an arbitrary y_{ij} for which $t - 1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$ have given values. This obvious modification results in Theorem 1.

4.2 *Proof of Theorem* 2. By Corollary 2, the maximum bias of the $[(1 + t_2) \cdots (1 + t_K)] \times [t_1 n]$ array is less than or equal to β_1. In general, $2 \leq w \leq K$, by Corollary 2 the maximum bias of the $[(1 + t_{w+1}) \cdots (1 + t_K)] \times [t_1 \cdots t_w n]$ array is less than or equal to β_w. Finally, by Corollary 1, if exactly $t - 1$ of $Y_{1h}, \cdots, Y_{(g-1)h}, Y_{(g+1)h}, \cdots, Y_{t_K h}$ have known values, $(1 \leq t \leq t_K)$, the maximum bias for the binary digit Y_{gh} is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t].$$

The inequality (3) is an immediate consequence of the relation

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^s/(\tfrac{1}{2} + \alpha)^s]/[1 + (\tfrac{1}{2} - \alpha)^s/(\tfrac{1}{2} + \alpha)^s] \leq 2s\alpha^2.$$

4.3 *Proof of Theorem* 3. From the given condition

$$t_K \geq 1/[Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1].$$

From this inequality for t_K it follows that

$$Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1 > 0.$$

Thus

$$t_{K-1} > 1/[Ct_1 \cdots t_{K-2}/(1 + t_1) \cdots (1 - t_{K-2}) - 1].$$

In general, $3 \leq w \leq K - 1$, given

$$t_w > 1/[Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1]$$

it follows that

$$Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1 > 0$$

whence

$$t_{w-1} > 1/[Ct_1 \cdots t_{w-2}/(1 + t_1) \cdots (1 + t_{w-2}) - 1].$$

Finally

$$t_1 > 1/(C - 1).$$

REFERENCES

[1] H. Burke Horton, "A method for obtaining random numbers," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 81–85.

[2] H. Burke Horton and R. Tynes Smith III, "A direct method for producing random digits in any number system," *Annals of Math. Stat*, Vol. 20 (1949), pp. 82–90.

THE DISTRIBUTION OF EXTREME VALUES IN SAMPLES WHOSE MEMBERS ARE SUBJECT TO A MARKOFF CHAIN CONDITION

By Benjamin Epstein

Department of Mathematics, Wayne University

1. Introduction. The extreme value problem as treated in the literature concerns itself with the following question: To find the distribution of the smallest, largest, or more generally the νth largest, or νth smallest values in random samples of size n, drawn from a distribution whose probability law is given by the d.f. $F(x)$. In this formulation the observed sample values $x_1, \cdots, x_n$ are assumed to be statistically independent. While the assumption of independence may be a good approximation to the true state of affairs in some cases, there are situations where this assumption is not justified.

Suppose, for instance, that the observations in the sample are ordered in time. Then it may happen that successive observations are stochastically dependent, the extent of this dependence being a function of the time interval separating these observations.[1] In such cases the present distribution theory for extreme values in samples of size n is inadequate and must be replaced by more general results.

It is clear that a clean-cut analytic solution to the problem of the distribution of extreme values in samples whose members may be stochastically dependent can be expected only for certain special kinds of dependence among successive observations. We are able, in this paper, to obtain the distribution of smallest, largest, second smallest, and second largest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process.

2. The distribution of smallest and largest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process. In this section the following assumption is made:

(A) observations $x_1, x_2, \cdots, x_n, \cdots$ are taken in order at times $t = 1$, $t = 2, \cdots, t = n, \cdots$ from a stationary Markoff random process.

The only information needed in the investigation of a stationary Markoff process at integral values of time is the function

$$F_2(x, y) = \text{Prob}\,(x_i \leq x, x_{i+1} \leq y), \tag{1}$$

independently of i, where $F_2(x, y)$ must be such that the marginal distribution obtained by integrating over x or y (if x_i or x_{i+1} take on a continuous range of

[1] If the observations $x_1, x_2, \cdots, x_n, \cdots$ are taken at discrete times $t_1, t_2, \cdots, t_n, \cdots$ a measure of stochastic dependence between x_i and x_j is the ordinary coefficient of correlation r_{ij}. If the observations are taken from a continuous stochastic process a natural measure of stochastic dependence between observations made at two different times is the covariance function of the process. In this paper we shall limit ourselves to processes which are discrete in time.

values) or summing over the possible values of x_i or x_{i+1} (if x_i and x_{i+1} can take on only discrete values) is of the form

$$F_1(x) = \text{Prob } (x_i \leq x), \tag{2}$$

independently of i.

An example of a random process meeting condition A is furnished by the Ornstein-Uhlenbeck process [1; 2]. In this case the joint d f. of x_i and x_{i+1} is given by a non-singular bivariate Gaussian distribution. The results in the present paper are stated completely in terms of the d.f.'s $F_2(x, y)$ and $F_1(x)$ defining the stationary Markoff process and will in particular be valid for observations taken at uniformly spaced time intervals from an Ornstein-Uhlenbeck process.

In this section we shall find the distribution of smallest and largest values in samples $x_1, x_2, \cdots, x_n$ drawn from a random process under assumption A and specified by the bivariate d.f. $F_2(x, y)$ and the associated one dimensional marginal d.f. $F_1(x)$. We first prove Theorem I.

THEOREM I. *Under assumption A, the distribution of largest values in samples of size n is given by the* d.f. $G_n^{(1)}(x) = [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2}$.

To prove this result we note that $G_n^{(1)}(x)$, the probability that the largest value in samples of size n is $\leq x$, is given by

$$G_n^{(1)}(x) = \text{Prob } (x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x). \tag{3}$$

To evaluate the right-hand side of (3) we proceed as follows:

$$\text{Prob } (x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) = \tag{4}$$
$$\text{Prob } (x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x) \text{ Prob } (x_n \leq x \mid x_1 \leq x, \cdots, x_{n-1} \leq x).$$

But under assumption A, (4) becomes

$$\text{Prob } (x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) = \tag{5}$$
$$\text{Prob } (x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x) \text{ Prob } (x_n \leq x \mid x_{n-1} \leq x)$$

or

$$G_n^{(1)}(x) = G_{n-1}^{(1)}(x) \text{ Prob } (x_n \leq x \mid x_{n-1} \leq x). \tag{5'}$$

But according to assumption A, and (1) and (2)

$$\text{Prob } (x_n \leq x \mid x_{n-1} \leq x) = \text{Prob } (x_{n-1} \leq x, x_n \leq x)/\text{Prob } (x_{n-1} \leq x) \tag{6}$$
$$= F_2(x, x)/F_1(x).$$

Therefore

$$G_n^{(1)}(x) = G_{n-1}^{(1)}(x)\, F_2(x, x)/F_1(x) \tag{7}$$
$$= G_1^{(1)}(x)\, (F_2(x, x))^{n-1}/(F_1(x))^{n-1}$$
$$= (F_2(x, x))^{n-1}/(F_1(x))^{n-2}.$$

This proves Theorem I.

For $n = 1, 2$, and 3 respectively one gets

$$(8) \quad G_1^{(1)}(x) = F_1(x), \quad G_2^{(1)}(x) = F_2(x, x), \quad G_3^{(1)}(x) = [F_2(x, x)]^2/F_1(x).$$

THEOREM II. *Under assumption A, the distribution of smallest values in samples of size n is given by the* d.f.

$$(9) \quad H_n^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^n}{[1 - F_1(x)]^{n-2}}.$$

To prove this result we first note that $H_n^{(1)}(x)$, the probability that the smallest value in samples of size n be $\leq x$ is given by,

$$1 - \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x).$$

To evaluate $H_n^{(1)}(x)$ we proceed as follows:

$$(10) \quad \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) =$$

$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)\,\text{Prob}\,(x_n > x \mid x_1 > x, \cdots, x_{n-1} > x).$$

But under assumption A, (10) becomes

$$(11) \quad \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) =$$

$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)\,\text{Prob}\,(x_n > x \mid x_{n-1} > x).$$

But

$$(12) \quad \text{Prob}\,(x_n > x \mid x_{n-1} > x) = \text{Prob}\,(x_{n-1} > x, x_n > x)/\text{Prob}\,(x_{n-1} > x).$$

To evaluate $\text{Prob}\,(x_{n-1} > x, x_n > x)$ we note that

$$(13) \quad \text{Prob}\,(x_{n-1} > x, x_n > x) + \text{Prob}\,(x_{n-1} \leq x, x_n > x)$$

$$+ \text{Prob}\,(x_{n-1} > x, x_n \leq x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x) = 1.$$

Also

$$(14) \quad \text{Prob}\,(x_{n-1} \leq x, x_n > x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x)$$

$$= \text{Prob}\,(x_{n-1} \leq x),$$

and

$$(15) \quad \text{Prob}\,(x_{n-1} > x, x_n \leq x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x)$$

$$= \text{Prob}\,(x_n \leq x).$$

Recalling that

$$(16) \quad F_2(x, x) = \text{Prob}\,(x_{n-1} \leq x, x_n \leq x)$$

and

$$(17) \quad F_1(x) = \text{Prob}\,(x_{n-1} \leq x) = \text{Prob}\,(x_n \leq x)$$

we get

$$\text{Prob}\,(x_{n-1} > x, x_n > x) = 1 - 2F_1(x) + F_2(x, x). \tag{18}$$

Therefore (10) becomes

$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x, x_n > x) = \tag{19}$$
$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)[1 - 2F_1(x) + F_2(x, x)]/(1 - F_1(x)).$$

Applying the recursion formula (19) successively we obtain

$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) = \tag{20}$$
$$\text{Prob}\,(x_1 > x)[1 - 2F_1(x) + F_2(x, x)]^{n-1}/[1 - F_1(x)]^{n-1}$$
$$= [1 - 2F_1(x) + F_2(x, x)]^{n-1}/[1 - F_1(x)]^{n-2}.$$

Therefore $H_n^{(1)}(x)$, the probability that the smallest value in samples of size n is $\leq x$, is given by:

$$H_n^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^{n-1}}{[1 - F_1(x)]^{n-2}}. \tag{21}$$

This completes the proof of Theorem II.

In particular for $n = 1$, 2, and 3 respectively the d.f.'s of the smallest value in samples of size n are given by:

$$H_1^{(1)}(x) = F_1(x), \qquad H_2^{(1)}(x) = 2F_1(x) - F_2(x, x),$$
$$H_3^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^2}{1 - F_1(x)}. \tag{22}$$

3. Distribution of the second largest and second smallest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process. Under assumption A of Section II we can state the following theorem.

THEOREM III. *Under assumption A the distribution of second largest values in samples of size n, $n \geq 2$, is given by the* d.f. $G^{(2)}(x)$,

$$G_n^{(2)}(x) = [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2}$$
$$+ 2[F_2(x, x)]^{n-2}\{F_1(x) - F_2(x, x)\}/[F_1(x)]^{n-2}$$
$$+ (n - 2)\,[F_2(x, x)]^{n-3}\,\{F_1(x) - F_2(x, x)\}^2/[F_1(x)]^{n-3}(1 - F_1(x)).$$

To prove this result we first note that $G_n^{(2)}(x)$, the probability that the second largest value is $\leq x$, is given by

$$G_n^{(2)}(x) = \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x)$$
$$+ \text{Prob}\,(x_1 > x, x_2 \leq x, x_3 \leq x, \cdots, x_n \leq x)$$
$$+ \text{Prob}\,(x_1 \leq x, x_2 > x, x_3 \leq x, x_4 \leq x, \cdots, x_n \leq x) + \cdots \tag{23}$$
$$+ \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-2} \leq x, x_{n-1} > x, x_n \leq x)$$
$$+ \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x, x_n > x).$$

According to Theorem I

$$\text{(24)} \qquad \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) = [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2}.$$

It can readily be shown that

$$\begin{aligned} &\text{Prob}\,(x_1 > x, x_2 \leq x, x_3 \leq x, \cdots, x_n \leq x) \\ \text{(25)} \qquad &= \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x, x_n > x) \\ &= [F_2(x, x)]^{n-2}\,[F_1(x) - F_2(x, x)]/[F_1(x)]^{n-2}. \end{aligned}$$

It can also be shown that each of the remaining $(n-2)$ terms on the right-hand side of (23) is equal to

$$\text{(26)} \qquad [F_2(x, x)]^{n-3}\,[F_1(x) - F_2(x, x)]^2/[F_1(x)]^{n-3}(1 - F_1(x)).$$

Combining (23), (24), (25), and (26) we get the desired result in Theorem III, i.e.,

$$\begin{aligned} G_n^{(2)}(x) &= [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2} \\ \text{(27)} \qquad &+ 2[F_2(x, x)]^{n-2}\,[F_1(x) - F_2(x, x)]/[F_1(x)]^{n-2} \\ &+ (n-2)[F_2(x, x)]^{n-3}\,[F_1(x) - F_2(x, x)]^2/[F_1(x)]^{n-3}(1 - F_1(x)). \end{aligned}$$

In a similar way one can prove Theorem IV.

THEOREM IV. *Under assumption A, the distribution of second smallest values in samples of size n, $n \geq 2$, is given by the* d.f. $H_n^{(2)}(x)$.

$$\begin{aligned} H_n^{(2)}(x) &= 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^{n-1}}{[1 - F_1(x)]^{n-2}} \\ \text{(28)} \qquad &- 2\,\frac{[1 - 2F_1(x) + F_2(x, x)]^{n-2}}{[1 - F_1(x)]^{n-2}}\,[F_1(x) - F_2(x, x)] \\ &- (n-2)\,\frac{[1 - 2F_1(x) + F_2(x, x)]^{n-3}}{[1 - F_1(x)]^{n-3}}\,\frac{[F_1(x) - F_2(x, x)]^2}{F_1(x)}. \end{aligned}$$

REFERENCES

[1] J. L. DOOB, "The brownian movement and stochastic equations," *Annals of Mathematics*, Vol. 43 (1942), pp. 351.

[2] M. C. WANG AND G. E. UHLENBECK, "On the theory of the brownian motion II," *Reviews of Modern Physics*, Vol. 17 (1945), p. 323.

NOTES

This section is devoted to brief research and expository articles and other short items.

NOTE ON THE CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE[1]

By Abraham Wald

Columbia University

1. Introduction. The problem of consistency of the maximum likelihood estimate has been treated in the literature by several authors (see, for example, Doob [1][2] and Cramér [2][3]). The purpose of this note is to give another proof of the consistency of the maximum likelihood estimate which may be of interest because of its relative simplicity and because of the easy verifiability of the underlying assumptions. The present proof has some common features with that given by Doob, insofar that both proofs make no differentiability assumptions (thus, not even the existence of the likelihood equation is postulated) and both are based on the strong law of large numbers and an inequality involving the log of a random variable. The assumptions in the present note are stronger in some respects than those made by Doob, but also the results obtained here are stronger. For the sake of simplicity, the author did not attempt to give the most general results or to weaken the underlying assumptions as much as possible. Remarks on possible generalizations are made in Section 4.

Let $X_1, X_2, \cdots$, etc. be independently and identically distributed chance variables. The most frequently considered case in the literature is that where the common distribution is known, except for the values of a finite number of

[1] The author wishes to thank J. L. Doob for several comments and suggestions he made in connection with this note.

[2] According to a communication from Doob, his Theorem 4 is incorrect, but is correct if the class of almost everywhere continuous functions in that theorem is replaced by a suitable class C of functions. The class C can be any one of a variety of classes; for example, the class of bounded almost everywhere continuous functions, or the larger class of almost everywhere continuous functions each of which is less than or equal in modulus to any one of a prescribed sequence of functions with finite expectations. His Theorem 5 on the consistency of the maximum likelihood is then dependent on the class C used in Theorem 4.

[3] The proof given by Cramér [2], pp. 500–504, establishes the consistency of some root of the likelihood equation but not necessarily that of the maximum likelihood estimate when the likelihood equation has several roots. Recently, Huzurbazar [3] showed that under certain regularity conditions the likelihood equation has at most one consistent solution and that the likelihood function has a relative maximum for such a solution. Since there may be several solutions for which the likelihood function has relative maxima, Cramér's and Huzurbazar's results taken together still do not imply that a solution of the likelihood equation which makes the likelihood function an absolute maximum is necessarily consistent.

parameters, $\theta^1, \theta^2, \cdots, \theta^k$. In this note we shall treat the parametric case. For any parameter point $\theta = (\theta^1, \cdots, \theta^k)$, let $F(x, \theta)$ denote the corresponding cumulative distribution function of X_i; i.e., $F(x, \theta) = \text{prob.}\{X_i < x\}$. The totality Ω of all possible parameter points is called the parameter space. Thus, the parameter space Ω is a subset of the k-dimensional Cartesian space.

It is assumed in this note that for any θ, the cumulative distribution function $F(x, \theta)$ admits an elementary probability law $f(x, \theta)$. If $F(x, \theta)$ is absolutely continuous, $f(x, \theta)$ denotes the density at x. If $F(x, \theta)$ is discrete, $f(x, \theta)$ is equal to the probability that $X_i = x$.

Throughout this note the following assumptions will be made.

ASSUMPTION 1. *$F(x, \theta)$ is either discrete for all θ or is absolutely continuous for all θ.*

Before formulating the next assumption, we shall introduce the following notations: for any θ and for any positive value ρ let $f(x, \theta, \rho)$ be the supremum of $f(x, \theta')$ with respect to θ' when $|\theta - \theta'| \leq \rho$. For any positive r, let $\varphi(x, r)$ be the supremum of $f(x, \theta)$ with respect to θ when $|\theta| > r$. Furthermore, let $f^*(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) > 1$, and $= 1$ otherwise. Similarly, let $\varphi^*(x, r) = \varphi(x, r)$ when $\varphi(x, r) > 1$, and $= 1$ otherwise.

ASSUMPTION 2. *For sufficiently small ρ and for sufficiently larger r the expected values $\int_{-\infty}^{\infty} \log f^*(x, \theta, \rho)\, dF(x, \theta_0)$ and $\int_{-\infty}^{\infty} \log \varphi^*(x, r)\, dF(x, \theta_0)$ are finite where θ_0 denotes the true parameter point.*[4]

ASSUMPTION 3. *If $\lim_{i=\infty} \theta_i = \theta$, then $\lim_{i=\infty} f(x, \theta_i) = f(x, \theta)$ for all x except perhaps on a set which may depend on the limit point θ (but not on the sequence θ_i) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point θ_0.*

ASSUMPTION 4. *If θ_1 is a parameter point different from the true parameter point θ_0, then $F(x, \theta_1) \neq F(x, \theta_0)$ for at least one value of x.*

ASSUMPTION 5. *If $\lim_{i=\infty} |\theta_i| = \infty$, then $\lim_{i=\infty} f(x, \theta_i) = 0$ for any x except perhaps on a fixed set (independent of the sequence θ_i) whose probability is zero according to the true parameter point θ_0.*

ASSUMPTION 6. *For the true parameter point θ_0 we have*

$$\int_{-\infty}^{\infty} |\log f(x, \theta_0)|\, dF(x, \theta_0) < \infty.$$

ASSUMPTION 7. *The parameter space Ω is a closed subset of the k-dimensional Cartesian space.*

ASSUMPTION 8. *$f(x, \theta, \rho)$ is a measurable function of x for any θ and ρ.*

It is of interest to note that if we forbid the dependence of the exceptional set on θ in Assumption 3, Assumption 8 is a consequence of Assumption 3, as can easily be verified.

---

[4] The measurability of the functions $f^*(x, \theta, \rho)$ and $\varphi^*(x, r)$ for any θ, ρ and r follows easily from Assumption 8.

In the discrete case, Assumption 8 is unnecessary. In fact, we may replace $f(x, \theta, \rho)$ everywhere by $\bar{f}(x, \theta, \rho)$ where $\bar{f}(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta_0) > 0$, and $\bar{f}(x, \theta, \rho) = 1$ when $f(x, \theta_0) = 0$ Here θ_0 denotes the true parameter point. Since $f(x, \theta_0) > 0$ only for countably many values of x, $\bar{f}(x, \theta, \rho)$ is obviously a measurable function of x.

In the absolutely continuous case, $F(x, \theta)$ does not determine $f(x, \theta)$ uniquely. If Assumptions 3, 5 and 8 hold for one choice of $f(x, \theta)$, they do not necessarily hold for another choice of $f(x, \theta)$. This is in a way undesirable, but assumptions of such nature are unavoidable if we want to insure the consistency of the maximum likelihood estimate. It is, however, possible to formulate assumptions which remain valid for all possible choices of $f(x, \theta)$ and which insure the consistency of the maximum likelihood estimate for a particular choice of $f(x, \theta)$ In this connection the following remark due to Doob is of interest. Let Assumptions 3′ and 5′ be the same as 3 and 5, respectively, except that the exceptional set is permitted to depend on the sequence θ_i. If 3′ and 5′ hold for one choice of $f(x, \theta)$, they also hold for any other choice. Doob has shown that Assumptions 3′ and 5′ insure the existence of a choice of $f(x, \theta)$ for which Assumptions 3, 5 and 8 hold. Thus, one may say that Assumptions 3′ and 5′ are the essential ones and the stronger assumptions 3, 5 and 8 are needed merely to exclude a "bad" choice of $f(x, \theta)$.

2. Some lemmas. In this section we shall prove some lemmas which will be used in the next section to obtain the main theorems. Let θ_0 be the true parameter point. By the expected value Eu of any chance variable u we shall mean the expected value determined under the assumption that θ_0 is the true parameter point. For any chance variable u, u' will denote the chance variable which is equal to u when $u > 0$ and equal to zero otherwise. Similarly, for any chance variable u, the symbol u'' will be used to denote the chance variable which is equal to u when $u < 0$ and equal to zero otherwise. We shall say that the expected value of u exists if $Eu' < \infty$. If the expected value of u' is finite but that of u'' is not, we shall say that the expected value of u is equal to $-\infty$.

LEMMA 1. *For any $\theta \neq \theta_0$ we have*

$$E \log f(X, \theta) < E \log f(X, \theta_0) \tag{1}$$

where X is a chance variable with the distribution $F(x, \theta_0)$.

PROOF. It follows from Assumption 2 that the expected values in (1) exist. Because of Assumption 6, we have

$$E \mid \log f(X, \theta_0) \mid < \infty. \tag{2}$$

If $E \log f(X, \theta) = -\infty$, Lemma 1 obviously holds. Thus, we shall merely consider the case when $E \log f(X, \theta) > -\infty$. Then

$$E \mid \log f(X, \theta) \mid < \infty. \tag{3}$$

Let $u = \log f(X, \theta) - \log f(X, \theta_0)$.[5] Clearly, $E \mid u \mid < \infty$. It is known that for

any chance variable u which is not equal to a constant (with probability one) and for which $E \mid u \mid < \infty$, we have[6]

$$Eu < \log Ee^u. \tag{4}$$

Since in our case

$$Ee^u \leqq 1, \tag{5}$$

and since u differs from zero on a set of positive probability (due to Assumption 4), we obtain from (4)

$$Eu < 0. \tag{6}$$

Thus, Lemma 1 is proved.

We shall now prove the following lemma.

LEMMA 2. $\lim_{\rho=0} E \log f(X, \theta, \rho) = E \log f(X, \theta)$.

PROOF. Let $f^*(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) \geqq 1$, and $= 1$ otherwise. Similarly, let $f^*(x, \theta) = f(x, \theta)$ when $f(x, \theta) \geqq 1$, and $= 1$ otherwise. It follows from Assumption 3 that

$$\lim_{\rho=0} \log f^*(x, \theta, \rho) = \log f^*(x, \theta) \tag{7}$$

except perhaps on a set whose probability measure is zero. Since $\log f^*(x, \theta, \rho)$ is an increasing function of ρ, it follows from (7) and Assumption 2 that

$$\lim_{\rho=0} E \log f^*(X, \theta, \rho) = E \log f^*(X, \theta). \tag{8}$$

Let $f^{**}(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) \leqq 1$, and $= 1$ otherwise. Similarly, let $f^{**}(x, \theta) = f(x, \theta)$ when $f(x, \theta) \leqq 1$, and $= 1$ otherwise. Clearly,

$$\mid \log f^{**}(x, \theta, \rho) \mid \leqq \mid \log f^{**}(x, \theta) \mid \tag{9}$$

and

$$\lim_{\rho=0} \log f^{**}(x, \theta, \rho) = \log f^{**}(x, \theta) \tag{10}$$

for all x except perhaps on a set whose probability measure is zero. The relation

$$\lim_{\rho=0} E \log f^{**}(X, \theta, \rho) = E \log f^{**}(X, \theta) \tag{11}$$

follows from (9) and (10) in both cases, when $E \log f^{**}(X, \theta)$ is finite and when $E \log f^{**}(X, \theta) = -\infty$. Lemma 2 is an immediate consequence of (8) and (11).

LEMMA 3. *The equation*

$$\lim_{r=\infty} E \log \varphi(X, r) = -\infty, \tag{12}$$

holds.

---

[5] It is of no consequence what value is assigned to u when $f(x, \theta)$ or $f(x, \theta_0)$ is zero, since the probability of such an event, because of (3), is zero.

[6] This is a generalization of the inequality between geometric and arithmetic means. See, for example, HARDY, LITTLEWOOD, POLYA, *Inequalities*, Cambridge 1934, p. 137, Theorem 184.

PROOF. It follows from Assumption 5 that

$$\lim_{r=\infty} \log \varphi(x, r) = -\infty, \tag{13}$$

for any x (except perhaps on a set of probability 0). Since according to Assumption 2,

$$E \log \varphi^*(X, r) < \infty, \tag{14}$$

and since $\log \varphi(x, r) - \log \varphi^*(x, r)$ and $\log \varphi^*(x, r)$ are decreasing functions of r, Lemma 3 follows easily from (13).

3. The main theorems. We shall now prove the following theorems.

THEOREM 1. *Let ω be any closed subset of the parameter space Ω which does not contain the true parameter point θ_0. Then*

$$\text{prob.} \left\{ \lim_{n=\infty} \frac{\operatorname{Sup}_{\theta \in \omega} f(X_1, \theta) f(X_2, \theta) \cdots f(X_n, \theta)}{f(X_1, \theta_0) f(X_2, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1. \tag{15}$$

PROOF. Let r_0 be a positive number chosen such that

$$E \log \varphi(X, r_0) < E \log f(X, \theta_0). \tag{16}$$

The existence of such a positive number follows from Lemma 3. Let ω_1 be the subset of ω consisting of all points θ of ω for which $|\theta| \leqq r_0$. With each point θ in ω_1 we associate a positive value ρ_θ such that

$$E \log f(X, \theta, \rho_\theta) < E \log f(X, \theta_0). \tag{17}$$

The existence of such a ρ_θ follows from Lemmas 1 and 2. Since the set ω_1 is compact, there exists a finite number of points $\theta_1, \cdots, \theta_h$ in ω_1 such that $S(\theta_1, \rho_{\theta_1}) + \cdots + S(\theta_h, \rho_{\theta_h})$ contains ω_1 as a subset. Here $S(\theta, \rho)$ denotes the sphere with center θ and radius ρ. Clearly,

$$0 \leqq \operatorname{Sup}_{\theta \in \omega} f(x_1, \theta) \cdots f(x_n, \theta) \leqq \sum_{i=1}^{h} f(x_1, \theta_i, \rho_{\theta_i}) \cdots f(x_n, \theta_i, \rho_{\theta_i}) + \varphi(x_1, r_0) \cdots \varphi(x_n, r_0).$$

Hence, Theorem 1 is proved if we can show that

$$\text{prob} \left\{ \lim_{n=\infty} \frac{f(X_1, \theta_i, \rho_{\theta_i}) \cdots f(X_n, \theta_i, \rho_{\theta_i})}{f(X_1, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1 \quad (i = 1, \cdots, h) \tag{18}$$

and

$$\text{prob} \left\{ \lim_{n=\infty} \frac{\varphi(X_1, r_0) \cdots \varphi(X_n, r_0)}{f(X_1, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1. \tag{19}$$

The above equations can be written as

$$\text{(20)}\qquad \text{prob}\left\{\lim_{n\to\infty}\sum_{\alpha=1}^{n}[\log f(X_\alpha,\theta_i,\rho_{\theta_i})-\log f(X_\alpha,\theta_0)]=-\infty\right\}=1 \qquad (i=1,\cdots,h)$$

and

$$\text{(21)}\qquad \text{prob}\left\{\lim_{n\to\infty}\sum_{\alpha=1}^{n}[\log\varphi(X_\alpha,r_0(-\log f(X_\alpha,\theta_0)]=-\infty\right\}=1.$$

These equations follow immediately from (16), (17) and the strong law of large numbers. This completes the proof of Theorem 1.

THEOREM 2. *Let $\bar\theta_n(x_1,\cdots,x_n)$ be a function of the observations $x_1,\cdots,x_n$ such that*

$$\text{(22)}\qquad \frac{f(x_1,\bar\theta_n)\cdots f(x_n,\bar\theta_n)}{f(x_1,\theta_0)\cdots f(x_n,\theta_0)}\geqq c>0 \text{ for all } n \text{ and for all } x_1,\cdots,x_n.$$

Then

$$\text{(23)}\qquad \text{prob}\{\lim_{n\to\infty}\bar\theta_n=\theta_0\}=1.$$

PROOF. It is sufficient to prove that for any $\epsilon>0$ the probability is one that all limit points $\bar\theta$ of the sequence $\{\bar\theta_n\}$ satisfy the inequality $|\bar\theta-\theta_0|\leqq\epsilon$. The event that there exists a limit point $\bar\theta$ of the sequence $\{\bar\theta_n\}$ such that $|\bar\theta-\theta_0|>\epsilon$ implies that $\operatorname{Sup}_{|\theta-\theta_0|\geqq\epsilon} f(x_1,\theta)\cdots f(x_n,\theta)\geqq f(x_1,\bar\theta_n)\cdots f(x_n,\bar\theta_n)$ for infinitely many n. But then

$$\text{(24)}\qquad \frac{\operatorname{Sup}_{|\theta-\theta_0|\geqq\epsilon} f(x_1,\theta)\cdots f(x_n,\theta)}{f(x_1,\theta_0)\cdots f(x_n,\theta_0)}\geqq c>0$$

for infinitely many n. Since, according to Theorem 1, this is an event with probability zero, we have shown that the probability is one that all limit points $\bar\theta$ of $\{\bar\theta_n\}$ satisfy the inequality $|\bar\theta-\theta_0|\leqq\epsilon$. This completes the proof of Theorem 2.

Since a maximum likelihood estimate $\hat\theta_n(x_1,\cdots,x_n)$, if it exists, obviously satisfies (22) with $c=1$, Theorem 2 establishes the consistency of $\hat\theta_n(x_1,\cdots,x_n)$ as an estimate of θ.

4. Remarks on possible generalizations. The method given in this note can be extended to establish the consistency of the maximum likelihood estimates for certain types of dependent chance variables for which the strong law of large numbers remains valid.

The assumption that the parameter space Ω is a subset of a finite dimensional Cartesian space is unnecessarily restrictive. Let Ω be any abstract space. All of

our results can easily be shown to remain valid if Assumptions 3, 5 and 7 are replaced by the following one:

ASSUMPTION 9. *It is possible to introduce a distance* $\delta(\theta_1, \theta_2)$ *in the space* Ω *such that the following four conditions hold:*

(i) *The distance* $\delta(\theta_1, \theta_2)$ *makes* Ω *to a metric space*

(ii) $\lim_{i=\infty} f(x, \theta_i) = f(x, \theta)$ *if* $\lim_{i=\infty} \theta_i = \theta$ *for any* x *except perhaps on a set which may depend on* θ *(but not on the sequence* θ_i*) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point* θ_0.

(iii) *If* θ_0 *is a fixed point in* Ω *and* $\lim_{i=\infty} \delta(\theta_i, \theta_0) = \infty$, *then* $\lim_{i=\infty} f(x, \theta_i) = 0$ *for any* x.

(iv) *Any closed and bounded subset of* Ω *is compact.*

REFERENCES

[1] J. L. DOOB, "Probability and statistics," *Trans Amer Math. Soc.*, Vol. 36 (1934).

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[3] V. S. HUZURBAZAR, "The likelihood equation, consistency and the maxima of the likelihood function," *Annals of Eugenics*, Vol. 14 (1948).

ON WALD'S PROOF OF THE CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE

BY J. WOLFOWITZ

Columbia University

This note is written by way of comment on the pretty and ingenious proof of the consistency of the maximum likelihood estimate which is due to Wald and is printed in the present issue of the *Annals*. The notation of this paper of Wald's will henceforth be assumed unless the contrary is specified.

The consistency of the maximum likelihood estimate is a "weak" rather than a "strong" property, in the technical meaning which these words have in the theory of probability, i.e., it is a property of distribution functions rather than of infinite sequences of observations. Prof. Wald actually proves strong convergence, which is more than consistency. His proof uses the strong law of large numbers, and he remarks that his method "can be extended to establish consistency of the maximum likelihood estimates for certain types of dependent chance variables for which the strong law of large numbers remains valid." Below we shall use Wald's lemmas to give a proof of consistency which employs only the weak law of large numbers. Not only does this proof have the advantage of being expeditious, but it can be extended to a larger class of dependent chance variables.

The consistency of the maximum likelihood estimate follows from the following

THEOREM. *Let* η *and* ϵ *be given, arbitrarily small, positive numbers. Let* $S(\theta_0, \eta)$ *be the open sphere with center* θ_0 *and radius* η, *and let* $\Omega(\eta) = \Omega - S(\theta_0, \eta)$. *Let*

Wald's Assumptions 1–4 hold. There exists a number $h(\eta)$, $0 < h < 1$, and another positive number $N(\eta, \epsilon)$ such that, for any $n > N(\eta, \epsilon)$,

$$P_0\left\{\frac{\sup_{\theta \in \Omega(\eta)} \prod_1^n f(X_i, \theta)}{\prod_1^n f(X_i, \theta_0)} \geq h^n\right\} < \epsilon$$

where P_0 is the probability of the relation in braces according to $f(x, \theta_0)$.

Proof: Proceed exactly as in the proof of Wald's Theorem 1 and obtain $r_0, \rho_1, \cdots, \rho_k$, so that the set theoretic sum of the open spheres $S(\theta_i, \rho_i)$, $i = 1, 2, \cdots, k$, covers the compact set which is the intersection of $\Omega(\eta)$ with the sphere $|\theta| \leq r_0$. Define $T(\theta_i)$, $i = 1, \cdots, k+1$, as follows:

$$-2T(\theta_i) = E \log f(X, \theta_i, \rho_i) - E \log f(X, \theta_0)$$

$(i = 1, \cdots, k)$

$$-2T(\theta_{k+1}) = E \log \varphi(X, r_0) - E \log f(X, \theta_0)$$

If any of the right members above are infinite let $T(\theta_i)$ be one, say. Thus all $T(\theta_i)$ are positive. Applying the weak law of large numbers we have that, for any i such that $1 \leq i \leq k+1$, there exists a positive number N_i such that, when $n > N_i$,

$$P_0\left\{\frac{\prod_1^n f(X_i, \theta_i, \rho_i)}{\prod_1^n f(X_i, \theta_0)} > \exp\{-nT(\theta_i)\}\right\} > \frac{\epsilon}{k+1}$$

$(i = 1, \cdots, k)$

$$P_0\left\{\frac{\prod_1^n \varphi(X_i, r_0)}{\prod_1^n f(X_i, \theta_0)} > \exp\{-nT(\theta_{k+1})\}\right\} > \frac{\epsilon}{k+1}.$$

From this the theorem follows immediately, with

$$N(\eta, \epsilon) = \max_i N_i$$

$$h(\eta) = \max_i \exp\{-T(\theta_i)\}.$$

The author is obliged to Prof. Wald for his kindness in making his paper available to the author.

A NOTE ON RANDOM WALK

By Herbert T. David

The Johns Hopkins University Institute for Cooperative Research

A random walk is defined as a series of discrete steps along the real line, here denoted by I. Each step is represented by the chance variable X, with sectionally continuous density function $f(x)$. The walk begins at any point a of I, and continues until a step carries us outside some subregion Ω of I. In this note, Ω is taken as a finite interval with upper bound D and lower bound $D - y$. The chance variables N and Z are, respectively, the number of steps required to end the walk, and the endpoint of the walk. The range of Z always excludes Ω

Below, we define $x = D - a$, and consider $E(N)$ as a function $G(x, y)$ of x and y. Under specified conditions, a differential equation (32) is derived, relating $G(0, y)$ and $G(x, y)$.

Let

$$\psi_1(t) = f(t - a) \tag{1}$$

$$\psi_n(t) = \int \cdots (n-1) \cdots \int \prod_{i=1}^{n-1} f(g_i) \, f\left(t - a - \sum_{i=1}^{n-1} g_i\right) dg_1 \cdots dg_{n-1}; \qquad n > 1 \tag{2}$$

where

$$\left[a + \sum_{i=1}^{i} g_i\right] \epsilon \Omega, \quad \text{for } i: 1, 2, \cdots, n-1.$$

Then

$$P\{Z \epsilon w_1, N = n\} = \int_{w_1} \psi_n(t)\, dt \qquad \text{for } w_1 \epsilon \bar{\Omega}$$

$$P\{Z \epsilon w_2, N = n\} = 0 \qquad \text{for } w_2 \epsilon \Omega.$$

Hence

$$P\{N = n\} = \int_{\bar{\Omega}} \psi_n(t)\, dt$$

$$E(N) = \sum_{i=1}^{\infty} \int_{\bar{\Omega}} \psi_i(t)\, dt. \tag{3}$$

The transformation $\{h_i = a + \sum_{j=1}^{i} g_j ; i : 1, \cdots, n-1\}$ gives for $\psi_n(t)$ the more convenient expression

$$\psi_n(t) = \int_a \cdots (n-1) \cdots \int_a f(h_1 - a) \cdot \prod_{i=2}^{n-1} f(h_i - h_{i-1}) f(t - h_{n-1})\, dh_1 \cdots dh_{n-1}. \tag{4}$$

The n-fold integral $\int_I \psi_n(t)\,dt$ is absolutely convergent, hence may be integrated first with respect to t. This gives, keeping the notation of (4)

$$\int_I \psi_n(t)\,dt = \int_0 \psi_{n-1}(h_{n-1})\,dh_{n-1}. \tag{5}$$

Assuming that $E(N)$ remains finite for all considered a and Ω, series (3) may be rearranged, giving: $E(N) = \sum_{i=1}^{\infty} B_i$ where

$$B_i = \sum_{j=i}^{\infty} \int_0 \psi_j(t)\,dt.$$

Now, $B_1 = \sum_{i=1}^{\infty} P\{N = i\} = 1$. Also, using (5) and induction on n, it is readily shown that $B_n = \int_0 \psi_{n-1}(t)\,dt$, so that

$$E(N) = 1 + \sum_{i=1}^{\infty} \int_0 \psi_i(t)\,dt \tag{6}$$

Define transformations $T_n : [g_i = D - h_i,\ i = 1, \cdots, n-1;\ g_n = D - t]$. Substituting expressions (1) and (4) in (6), transform the jth term of the summation by T_j. This gives

$$E(N) = 1 + \sum_{n=1}^{\infty} \int_0^{\nu} \cdots (n) \cdots \int_0^{\nu} f(x - g_1) \prod_{i=1}^{n-1} f(g_i - g_{i+1})\,dg_1 \cdots dg_n \tag{7}$$

where $x = D - a$.

By (7), $E(N)$ is a function of x and y; hence we write $E(N) = G(x, y)$.

Define:

$M(k)$: Max $f(t)$ for $|t| \leq k$.

K : Any number satisfying $K \leq [1 - \epsilon]/M(K)$.

R : Any region $[-\infty < x < \infty;\ 0 \leq y \leq K]$.

M : Max $f(t)$.

L : Any number satisfying $L \leq [1 - \epsilon]/M$.

R' : Any region $[-\infty < x < \infty;\ 0 \leq y \leq L]$.

In the ensuing argument, we shall assume that

$$(x, y) \in R. \tag{8}$$

This condition restricts certain one-dimensional and two-dimensional variables to regions over which some infinite series are uniformly convergent with respect to these variables. Uniform convergence is required to validate term-by-term differentiations and integrations, and to establish the continuity in one or two variables of certain functions represented by series.

Arguments dealing with the solution of integral equations (17), (20) and (25) are valid only under the more restrictive condition

(9) $$(x, y) \in R'$$

this being the general sufficiency condition for the existence of solutions. However, (17) and (20) enter the argument with respect only to the derivation of equation (21) which could have been derived, though in a more cumbersome manner, by a term by term comparison of the series expressions for $[\lambda_{01}(x, y)]$ $[G(y, y)]$ and for $[G_{01}(x, y)]$ $[\lambda(y, y)]$, this latter approach being valid under (8) Similarly, (25) is used only in obtaining (27), which could have been obtained by a direct manipulation of the series expression for $G(x, y)$, this approach also being valid under (8). Hence, all subsequent derivations hold, as long as $(x, y) \in R$

By (8), we may interchange summation and integration with respect to g in (7). This gives

(10) $$G(x, y) = 1 + \int_0^y f(x - g)G(g, y)\, dg.$$

(11) Assume that $f(t)$ has a continuous derivative everywhere

Then $f(t)$ is continuous and $G(x, y)$ is continuous by (7) and (8). Hence

(12) $f(x - g)G(g, y)$ and $d/dx\, f(x - g)G(g, y)$ are continuous in (x, g)

(13) $f(x - g)G(g, y)$ is continuous in (g, y).

Let $G_{i,}(x, y)$ denote

$$\frac{d^i}{dx^i}\frac{d^j}{dy^j} G(x, y).$$

Then, by (12), we may differentiate (10) with respect to x, and, since $f_{10}(x - g) = -f_{01}(x - g)$, an integration by parts yields

(14) $$G_{10}(x, y) = f(x)G(0, y) - f(x - y)G(y, y) + \int_0^y f(x - g)G_{10}(g, y)\, dg.$$

Further, under (8), $G_{01}(x, y)$ may be obtained by differentiating (7) term by term, and is continuous in (x, y). Hence, $f(x - g)G_{01}(g, y)$ is continuous in (g, y), and we may differentiate (10) with respect to y, giving

(15) $$G_{01}(x, y) = f(x - y)G(y, y) + \int_0^y f(x - g)G_{01}(g, y)\, dg.$$

Adding (14) to (15), dividing by $G(0, y)$ which is always greater or equal to 1, and letting

(16) $$\lambda(x, y) = [G_{10}(x, y) + G_{01}(x, y)]/G(0, y)$$

we obtain

(17) $$\lambda(x, y) = f(x) + \int_0^y f(x - g)\lambda(g, y)\, dg.$$

Under (9), (17) defines a function

$$\lambda(x, y) = f(x) + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(x - g_1) \cdot \prod_{i=1}^{n-1} f(g_i - g_{i+1})\, f(g_n)\, dg_1 \cdots dg_n. \tag{18}$$

By (8), this function is continuous in (x, y) and may be differentiated term by term with respect to y. Further, $\lambda_{01}(x, y)$ thus gotten is continuous in (x, y), so that $f(x - g)\lambda_{01}(g, y)$ is continuous in (g, y). Hence, (17) may be differentiated with respect to y, giving

$$\lambda_{01}(x, y) = f(x - y)\lambda(y, y) + \int_0^y f(x - g)\lambda_{01}(g, y)\, dg. \tag{19}$$

Since, under (9), the integral equation

$$\alpha(x, y) = f(x - y) + \int_0^y f(x - g)\alpha(g, y)\, dg \tag{20}$$

has a unique continuous solution for every fixed y, (15) and (19) give

$$\frac{\lambda_{01}(x, y)}{\lambda(y, y)} = \frac{G_{01}(x, y)}{G(y, y)}. \tag{21}$$

Hence

$$\int_0^y \frac{\lambda_{01}(x, y)\, dx}{\lambda(y, y)} = \int_0^y \frac{G_{01}(x, y)\, dx}{G(y, y)}$$

and

$$\frac{\dfrac{d}{dy}\displaystyle\int_0^y \lambda(x, y)\, dx}{\lambda(y, y)} = \frac{\dfrac{d}{dy}\displaystyle\int_0^y G(x, y)\, dx}{G(y, y)}. \tag{22}$$

$$\text{Let } f(t) = f(-t). \tag{23}$$

Then it is obvious from the definition that

$$G(0, y) = G(y, y). \tag{24}$$

Further, by (15),

$$\frac{G_{01}(x, y)}{G(y, y)} = f(x - y) + \int_0^y f(x - g)\, \frac{G_{01}(g, y)}{G(y, y)}\, dg \tag{25}$$

so that, under (9), (25) gives for $G_{01}(x, y)/G(y, y)$ the unique expression

$$f(x - y) + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(x - g_1) \prod_{i=1}^{n-1} f(g_i - g_{i+1}) f(g_n - y)\, dg_1 \cdots dg_n$$

which, by (23), is equal to

$$f(y-x)+\sum_{n=1}^{\infty}\int_0^{\nu}\cdots(n)\cdots\int_0^{\nu}f(y-g_n)\prod_{i=1}^{n-1}f(g_{i+1}-g_i)f(g_1-x)\,dg_1\cdots dg_n.$$

Since, under (8), we may interchange summation and integration with respect to x, it follows that

$$\int_0^{\nu}\frac{G_{01}(x,y)}{G(y,y)}\,dx=\int_0^{\nu}f(y-x)\,dx+\sum_{n=1}^{\infty}\int_0^{\nu}\cdots(n+1)\cdots \cdot\int_0^{\nu}f(y-g_n)\prod_{i=1}^{n-1}f(g_{i+1}-g_i)f(g_1-x)\,dg_1\cdots dg_n\,dx \tag{26}$$

which, by a change of integration indices and a referral to (7), is seen to equal $[G(y,y)-1]$. (26) thus gives

$$\int_0^{\nu}G_{01}(x,y)\,dx=G(y,y)[G(y,y)-1]. \tag{27}$$

Further, by (16), (24), and (27),

$$\int_0^{\nu}\lambda(x,y)\,dx=G(0,y)-1 \tag{28}$$

so that

$$\frac{d}{dy}\int_0^{\nu}\lambda(x,y)\,dx=\frac{d}{dy}G(0,y) \tag{29}$$

while (24) and (27) also yield

$$\frac{d}{dy}\int_0^{\nu}G(x,y)\,dx=[G(0,y)]^2. \tag{30}$$

Hence, by (22), (29), and (30),

$$\lambda(y,y)=\frac{d}{d_y}G(0,y)/G(0,y). \tag{31}$$

Finally, substituting (31) in (21), and remembering the definition of λ given in (16), we get, using (24),

$$G(0,y)[G_{11}(x,y)+G_{02}(x,y)]=\frac{d}{dy}G(0,y)[G_{10}(x,y)+2G_{01}(x,y)]. \tag{32}$$

The conditions under which (32) holds are, in summary, (8), (11), and (23). If $f(t)$ has an expansion

$$f(t)=\sum_{i=0}^{\infty}A_i t^i;\qquad |t|<T \tag{33}$$

it is clear from (7) that

$$G(x,y)=\sum_{i,j=0}^{\infty}B_{ij}x^iy^j \tag{34}$$

for $(x,y)\in S$, where $S:[T_0\le x\le T_1;\ 0\le y\le T_1+T_0];\ T_0\le 0,\ T_1<T$.

Substituting (34) in (32), and equating coefficients of like powers of (x, y), we obtain the recursion formulae

$$(35)\quad \sum_{j+k=n} B_{i,j}B_{i,k}[j][2k-j+1]! = \sum_{j+k=n-1} B_{i+1,j}B_{i,k}[j+1][j-k]; \quad i=0,1,\cdots,$$

From (10), it is readily verified that $B_{0,0} \neq 0$ for $i \neq 0$, so that equations (35) give solutions for the $B_{i,j}$ in terms of the $B_{i,0}$. These solutions are of interest since they show a one-to-one correspondence between the functions $G(0, y)$ and $G(x, y)$, for $(x, y) \in [R \cap S]$.

NUMERICAL INTEGRATION FOR LINEAR SUMS OF EXPONENTIAL FUNCTIONS

By Robert E. Greenwood

The University of Texas and the Institute for Numerical Analysis[1]

1. Introduction. The methods of numerical integration going by the names trapezoidal rule, Simpson's rule, Weddle's rule, and the Newton-Cotes formulae are of the type

$$(1)\quad \int_{-1}^{1} f(x)\,dx \approx \sum_{i=0}^{n} \lambda_{in} f(x_{in})$$

where the abscissae $\{x_{in}\}$ are uniformly distributed on a finite interval, chosen as $(-1, 1)$ for convenience,

$$(2)\quad x_{in} = -1 + \frac{2i}{n}, \qquad i = 0, 1, 2, \cdots, n,$$

and where the set of constants $\{\lambda_{in}\}$ depend on the name of the rule and the value of n but not on the function $f(x)$. Throughout this note all abscissae will be assumed to be uniformly distributed on $(-1, 1)$ unless the contrary is explicitly stated.

Since correspondence relation (1) involves $(n + 1)$ constants $\{\lambda_{in}\}$, it might be possible to choose $(n + 1)$ arbitrary functions $g_j(x)$, $j = 0, 1, 2, \cdots, n$, and require that the set $\{\lambda_{in}\}$ be the solution, if such exists, of the $(n + 1)$ simultaneous linear equations

$$(3)\quad \int_{-1}^{1} g_j(x)\,dx = \sum_{i=0}^{n} \lambda_{in} g_j(x_{in}), \qquad j = 0, 1, 2, \cdots, n.$$

Indeed, the selection

$$(4)\quad g_j(x) = x^j, \qquad j = 0, 1, 2, \cdots, n,$$

will give a set of $(n + 1)$ simultaneous equations of form (3) and the solution $\{\lambda_{in}\}$ is the set of Newton-Cotes weights for that value of n. The numerical evaluation

[1] This work was performed with the financial support of the Office of Naval Research of the Navy Department.

of $\{\lambda_{in}\}$ is best accomplished by other and more sophisticated methods, however.[2]

Because of linearity in both the integral and the finite summation, once the constants $\{\lambda_{in}\}$ have been determined for a specific set of functions $\{g_j(x)\}$, correspondence relation (1) is exact for any linear combination of that fundamental set. Thus, for example, for the fundamental set (4), correspondence relation (1) with the appropriate values $\{\lambda_{in}\}$ is exact for all polynomials of degree less than or equal to n.

Although tradition favors the set of functions (4), there is nothing compelling about such a selection. Indeed, two other possible choices might be

$$g_j(x) = e^{jx}, \qquad j = 0, 1, 2, \cdots, n, \tag{5}$$

and

$$g_j(x) = e^{jx}, \tag{6}$$

$$j = -m, -m+1, \cdots, 0, 1, \cdots, m-1, m; n = 2m.$$

These choices would seem to be appropriate whenever numerical methods are being applied to exponential growth curves or exponential decay curves.

2. Use of the basic set $g_j(x) = e^{jx}$. If integration relation (1) be made exact for the set $\{e^{jx}\}$, $j = 0, 1, \cdots, n$ with evenly spaced x abscissae, the set (3) of $(n+1)$ simultaneous linear equations in the unknowns $\{\lambda_{in}\}$, $i = 0, 1, \cdots, n$ is obtained. Call the solution of this system $\{a_{in}\}$, solution values for $n = 1, 2, 3, 4, 5, 6$ are tabulated below.

For the symmetric case where integration relation (1) is made exact for $\{e^{jx}\}$, $j = -m, -m+1, \cdots, m-1, m; n = 2m$, a similar but different set of linear equations (3) results for the unknowns $\{\lambda_{in}\}$. Call the solution of this system $\{b_{in}\}$. As implied above, only even values of n are used in order to preserve the symmetry, and values of $\{b_{in}\}$ are tabulated below for $n = 2, 4, 6$.

$n = 1$,	$a_{01} = 1\ 31303\ 5285$		
	$a_{11} = 0\ 68696\ 4715$		
$n = 2$,	$a_{02} = 0\ 21805\ 032^{+}$	$b_{02} = 0.32260\ 623^{-}$	
	$a_{12} = 1.49780\ 742$	$b_{12} = 1\ 35478\ 755$	
	$a_{22} = 0.28414\ 226^{-}$	$b_{22} = 0.32260\ 623^{-}$	
$n = 3$,	$a_{03} = 0.51324\ 284$		
	$a_{13} = 0.22445\ 055$		
	$a_{23} = 1\ 08155\ 527$		
	$a_{33} = 0.18075\ 134$		
$n = 4$,	$a_{04} = -0.13716\ 639^{+}$	$b_{04} = 0.15048\ 171$	
	$a_{14} = 1.40098\ 548$	$b_{14} = 0.73243\ 318$	

[2] Whittaker and Robinson, *The Calculus of Observations*, 4th Edition, (1946), London, pp. 152–156.

	$a_{24} = -0.30895$	914		$b_{24} = 0.23417$	022
	$a_{34} = 0.91710$	903		$b_{34} = 0.73243$	318
	$a_{44} = 0.12803$	103		$b_{44} = 0.15048$	171
$n = 5,$	$a_{05} = 0.68919$	3			
	$a_{15} = -1.07644$	3			
	$a_{25} = 2.12534$	6			
	$a_{35} = -0.63595$	6			
	$a_{45} = 0.79933$	8			
	$a_{55} = 0.09852$	18			
$n = 6,$	$a_{06} = -0.83607$			$b_{06} = 0.09443$	5
	$a_{16} = 3.54128$			$b_{16} = 0.53464$	7
	$a_{26} = -3.88102$			$b_{26} = 0.01139$	3
	$a_{36} = 3.32254$			$b_{36} = 0.71905$	0
	$a_{46} = -0.94685$			$b_{46} = 0.01139$	3
	$a_{56} = 0.72075$			$b_{56} = 0.53464$	7
	$a_{66} = 0.07937$	5^+		$b_{66} = 0.09443$	5

The computing service of the Institute for Numerical Analysis has supplied the author with most of the coefficients tabulated above.

3. Estimates of the error term. The choices of the coefficients $\{a_{in}\}$ and $\{b_{in}\}$ are such that integration relation (1) is exact whenever

$$f(x) = A_0 + A_1 e^x + \cdots + A_n e^{nx} \quad \text{and} \quad \lambda_{in} = a_{in}, \tag{7}$$

and whenever

$$f(x) = B_{-m}e^{-mx} + B_{-m+1}e^{-(m-1)x} + \cdots + B_0 + \cdots + B_m e^{mx} \text{ and } \lambda_{in} = b_{in}. \tag{8}$$

When $f(x)$ is not of these prescribed forms, the error in using correspondence (1) may be of some importance. By making the transformation

$$u = e^x, \qquad f(x) = f(\log u) = g(u) \tag{9}$$

integration relation (1) becomes

$$\int_{e^{-1}}^{e} g(u)\frac{du}{u} \simeq \sum_{i=0}^{n} \lambda_{in} g(u_{in}) \tag{10}$$

where the $\{u_{in}\}$ are not evenly distributed. By approximating $g(u)$ by its Taylor's series with a remainder term, the following expressions for the error in using correspondence (1) can be obtained:
Using the coefficients $\{a_{in}\}$,

$$\text{Error} \leq \frac{\left(\frac{e^2-1}{2e}\right)^{n+1}}{(n+1)!}\left[2 + \sum_{i=0}^{n} |a_{in}|\right]\left[\max_{-1\leq x\leq 1}\left(e^{-x}\frac{d}{dx}\right)^{n+1} f(x)\right] \tag{11}$$

and, using the coefficients $\{b_{in}\}$,

$$\text{Error} \leq \frac{\left(\frac{e^2-1}{2e}\right)^{2m+1}}{(2m+1)!}\left[\frac{e^m-e^{-m}}{m}+\sum_{i=0}^{2m}\frac{|b_{i,2m}|}{e^{mx_{i,2m}}}\right] \cdot \left[\max_{-1\leq x\leq 1}\left(e^{-x}\frac{d}{dx}\right)^{2m+1} e^{mx}f(x)\right]. \tag{12}$$

Neither of these error expressions can be said to be very practical in actual computation, and neither appears suitable for establishing convergence properties of the type

$$\lim_{n\to\infty}\sum_{i=0}^{n}\lambda_{in}f(x_{in}) = \int_{-1}^{1} f(x)\,dx. \tag{13}$$

However, both (11) and (12) reduce to zero when $f(x)$ is of the form prescribed by (7) or (8) respectively.

4. Numerical examples. As illustrative numerical examples, the case $n = 4$ was selected and several typical functions were integrated approximately by the positive power exponential rule, the symmetrical exponential rule and the Newton-Cotes formula,

$$\int_{-1}^{1} f(x)\,dx = \tfrac{1}{45}[7f(-1) + 32f(-\tfrac{1}{2}) + 12f(0) + 32f(\tfrac{1}{2}) + 7f(1)].$$

Values of $\{a_{i4}\}$ and $\{b_{i4}\}$ are given in the tables in part 2. The typical functions used were x^2, e^{2x}, $1/(x+3)$, e^{-x^2}, xe^x, x^6, and $e^{2.2x}$. The following results were obtained:

Function	*Positive Power Exponential*	*Symmetrical Exponential*	*Newton-Cotes*	*8 Decimal Approximation to Exact Value*
x^2	.5703 8827	.6671 8001	.6666 6666	.6666 6667
e^{2x}	3.6268 6044	3.6268 6041	3.6317 3108	3.6268 6041
$1/(x+3)$	.6828 6353	.6931 5792	.6931 7460	.6931 4718
e^{-x^2}	1.4930 1396	1.4857 2754	1.4887 4582	1.4936 4827
xe^x	.7292 4338	.7353 6007	.7361 7480	.7357 5888
x^6	.0270 8487	.3238 5190	.3333 3332	.2857 1429
$e^{2.2x}$	4.0527 7287	4.0530 7585	4.0607 7415	4.0519 1379

From this tabulation, it would appear that the symmetrical exponential method compares favorably with the Newton-Cotes method for such typical functions as $1/(x+3)$, e^{-x^2}, xe^x, x^6, and $e^{2.2x}$. Note that the choice of x^2 or e^{2x} is not really a fair choice when comparing these two methods, since Newton-Cotes is derived so as to give exactness for x^2 and the symmetrical exponential so as to give exactness for e^{2x}.

SMOOTHEST APPROXIMATION FORMULAS

By Arthur Sard[1]

Queens College

Introduction. Consider a process of approximation which operates on a function $x = x(t)$. The error in the process may be thought of as a sum $R + \delta A$, where R is the error that would be present if x were exact and δA is the error due to errors in x. (Precise definitions are given below.) Suppose that one wishes to choose one process A from a class $\mathfrak{A}$ of processes. In some situations it is appropriate to base the choice on R alone[2]; in others it is appropriate to consider δA.

The primary purpose of the present note is to formulate a criterion of smoothest approximation: That A in $\mathfrak{A}$ is smoothest which minimizes the variance of δA. A criterion based on both R and δA is also suggested. (Sections 1 and 2.) Smoothest approximate integration formulas of one type are derived in Section 3.

Progress in the technique of estimating the covariance function of the errors in x will lead to further applications of the criterion of smoothest approximation.

1. Approximation of a functional. Suppose that X is a space of functions $x = x(t)$ each of which is continuous on $a \leqq t \leqq b$. Let $f[x]$ be a functional defined on X; that is, $f[x]$ is a real number defined for each $x \in X$. For example, X might be the space of functions with second derivatives on $[a, b]$ and $f[x]$ might be $x''(u)$, where u is a fixed number in $[a, b]$.

Suppose that $f[x]$ is to be approximated by a Stieltjes integral

$$A = \int_a^b x(t)\, d\alpha(t), \qquad x \in X, \tag{1}$$

where α is a function of bounded variation. The remainder in the approximation of $f[x]$ by A is

$$R = A - f[x].$$

If the approximation (1) operates on $x + \delta x$ instead of x, the result is $A + \delta A = \int_{t=a}^{b} (x + \delta x)\, d\alpha$; and the error in the approximation of $f[x]$ by $A + \delta A$ is $R + \delta A$, where

$$\delta A = \int_a^b \delta x(t)\, d\alpha(t). \tag{2}$$

Consider a class $\mathfrak{A}$ of approximations A, each of the form (1). We shall propose a criterion for characterizing the "smoothest A" in $\mathfrak{A}$, relative to the covariance function of the errors δx.

---

[1] The author gratefully acknowledges financial support received from the Office of Naval Research.

[2] "Best approximate integration formulas; best approximation formulas," *Amer. Jour. of Math.*, Vol. 71 (1949), pp. 80–91.

Assume that $\delta x = \delta x(t)$ is a stochastic process with mean zero[3] and covariance function $\sigma(t, u) = E[\delta x(t)\delta x(u)]$. Then, by (2), δA is a stochastic variable; and[4]

$$E\delta A = E\int_{t=a}^{b} \delta x \, d\alpha = \int_{t=a}^{b} 0 \, d\alpha = 0,$$

$$E(\delta A)^2 = v = E\left[\int_a^b \delta x(t)\, d\alpha(t) \int_a^b \delta x(u)\, d\alpha(u)\right] = \int_a^b \int_a^b \sigma(t,u)\, d\alpha(t)\, d\alpha(u). \tag{3}$$

CRITERION. *That A (if any) in $\mathfrak{A}$ is smoothest which minimizes the variance v of δA.*

In particular cases, this criterion (least squares) has been proposed and used by Chebyshev and others. An application to approximate integration is given in section 3 below.

One may extend this discussion to cases in which the approximations A involve derivatives of x.

Remark. The criterion of *best* approximation[2] may be combined with the above criterion of *smoothest* approximation as follows: That A (if any) in $\mathfrak{A}$ is the best compromise which minimizes a specified combination of the variance of δA and the modulus of R. Here it is assumed that the remainders R satisfy the conditions for the existence of the modulus.[2]

2. Approximation of a function. One may extend the preceding discussion to the case in which $y = f[x]$ is an operation to a space of functions $y = y(u)$, $a \leqq u \leqq b$; and in which the approximation of $f[x]$ is

$$A = \int_a^b x(t)\, d_t \alpha(t, u), \qquad x \in X,$$

where, for each u, α is a function of bounded variation in t. Then, for each u, δA has a variance $v(u)$. *Criterion.* That A (if any) in a class of approximations is smoothest which minimizes $v(u)$ for all u; failing such an A, that A (if any) is smoothest which minimizes the integral of $v(u)$, or alternatively, the supremum of $v(u)$, over $a \leqq u \leqq b$.

3. Smoothest approximate integration formulas in a particular case.[5] Let m and n be fixed integers; $m \geqq 1$, $n \geqq 0$. *Let $\mathfrak{A} = \mathfrak{A}_{m,n}$ be the class of all approximations of*

[3] The essential point here is that $E\delta(t) = m(t)$ be known for each t, for given $m(t)$, one could and would replace $x \to \delta x$ by $x \to \delta x - m$.

[4] We assume here that the integrals in (3) exist and that the inversions of E and $\int d\alpha$ are valid. For this it is sufficient that δx be integrable relative to the product measure $\alpha\omega$ for all functions α corresponding to elements of $\mathfrak{A}$, where ω is the measure in the underlying probability space relative to which E is the operator $\int d\omega$. Cf. J. L. Doob, "Probability in function space," *Bull. Amer. Math. Soc.*, Vol. 53 (1947), especially pp. 26, 27.

[5] The approximate integration formulas of this section are of such a nature that one would expect them to be known. The values of J at the end are probably new

$$\int_{-m/2}^{m/2} x(t)\,dt = f[x]$$

of the form

$$A = \sum_{i=-m/2}^{m/2} b_i x(i),$$

the $m + 1$ *constants* b_i *being such that* $A = f[x]$ *whenever* $x(t)$ *is a polynomial of degree* n. Throughout this section i is to range over the $m + 1$ values $i = -m/2$, $-m/2 + 1, \cdots, +m/2$. *Suppose that the errors* $\delta x(i)$ *are independent, with common variance* σ^2, *and with mean zero.* Then $\alpha(t)$ is a step function with jumps b_i at $t = i$; and

$$v = \sigma^2 \sum_i b_i^2.$$

The smoothest approximation in $\mathfrak{C}_{m,n}$ is the one for which v is a minimum. (The $m + 1$ variables b_i in v are subject to $n + 1$ constraints due to the condition that the approximation be exact for degree n. The set $\mathfrak{C}_{m,n}$ is empty if and only if m is less than the largest even integer contained in n.)

If $n = 0$ or 1, the smoothest formula in $\mathfrak{C}_{m,n}$ is the one for which all the coefficients are equal:

$$b_i = m/(m + 1);$$

in which case

$$v = m^2\sigma^2/(m + 1).$$

If $n = 2$ or 3, the smoothest formula in $\mathfrak{C}_{m,n}$ is characterized by the following relations:

$$b_i = \lambda_0 + i^2\lambda_1,$$

$$\lambda_0 = m(2m^2 + 9m - 6)/2(m - 1)(m + 1)(m + 3),$$

$$\lambda_1 = -30m/(m - 1)(m + 1)(m + 2)(m + 3);$$

in which case

$$v/\sigma^2 = \lambda_0 m + \lambda_1 m^3/12.$$

Thus, the smoothest approximation in $\mathfrak{C}_{3,2}$ or in $\mathfrak{C}_{3,3}$ is the following:

$$A = \tfrac{1}{2}[x(-3) + x(3)] + \tfrac{9}{14}[x(-2) + x(2)] + \tfrac{1}{14}[x(-1) + x(1)] + \tfrac{4}{7}\,x(0).$$

By the method of Lagrange's multipliers, one may establish the following relations for the smoothest formula in $\mathfrak{C}_{m,n}$. Here i has the same range of values as before; μ and ν range over $0, 1, \cdots, [n/2]$.

$$b_i = \sum_\mu \lambda_\mu i^{2\mu},$$

$$v/\sigma^2 = \sum_\mu \lambda_\mu c_\mu,$$

where

$$c_\mu = m^{2\mu+1}/4^\mu(2\mu+1),$$

and λ_μ are determined by the equations

$$\sum_\mu \lambda_\mu \sum_i i^{2(\mu+\nu)} = c_\nu .$$

The class $\mathfrak{A}_{m,n}$ is such that for each $A \in \mathfrak{A}_{m,n}$ there is a function $k(t)$ with the following property:[2]

$$R = A - f[x] = \int_{-m/2}^{m/2} x^{(n+1)}(t)k(t)\,dt,$$

whenever x is a function with continuous $(n+1)$th derivative. The quantity

$$J = \int_{-m/2}^{m/2} k^2(t)\,dt$$

is useful in appraising R, since

$$R^2 \leqq J \int_{-m/2}^{m/2} x^{(n+1)}(t)^2\,dt,$$

by Schwarz's inequality.

Values of J for the smoothest formulas are as follows.

$$n = 0 : J = m^3/6(m+1).$$

$$n = 1 : J = m^3(3m^2+2m+1)/360(m+1).$$

For $n = 2$ and 3, and $m \leqq 6$, the numerical values of J are as follows.

m	J $(n = 2)$.	J $(n = 3)$.
2	1/1,890	1/9,072
3	11/8,960	13/17,920
4	134/33,075	62,539/13,891,500
5	1,865/150,528	136,223/6,322,176
6	8/245	6,683/82,320

For the method of calculation of J, as well as the transformation of J under a linear transformation of t, the reader may consult the paper[2].

ON THE POWER FUNCTION OF THE "BEST" t-TEST SOLUTION OF THE BEHRENS-FISHER PROBLEM

By John E. Walsh

The Rand Corporation

1. Introduction. The Behrens-Fisher problem is concerned with significance tests for the difference of the means of two normal populations when the ratio of the variances of the populations is unknown. Denote one population by $N(a_1, \sigma_1^2)$ and the other by $N(a_2, \sigma_2^2)$, where the notation $N(a, \sigma^2)$ represents a normal population with mean a and variance σ^2. Let m sample values be drawn from $N(a_1, \sigma_1^2)$ and n sample values from $N(a_2, \sigma_2^2)$ where $m \leq n$. Then Scheffé [1] has shown that certain optimum properties are possessed by a t-test solution he proposed for the Behrens-Fisher problem, in which the numerator of t is based on the difference of the means of the samples while the denominator is based on the square root of a function of the sample values which has a χ^2-distribution with $m - 1$ degrees of freedom. The purpose of this note is to compare the power function of this t-test with the power function of the corresponding most powerful test for the case in which the ratio of variances σ_1^2/σ_2^2 is also known (only one-sided and symmetrical tests are considered). This comparison is made by computing the power efficiency (see section 2 for definition) of Scheffé's test.

It is sufficient to limit power efficiency investigations to one-sided tests. As shown in [2], a symmetrical t-test with significance level 2α has the same power efficiency as the corresponding one-sided t-test with significance level α. Equation (2) of section 2 furnishes an explicit formula whereby approximate power efficiencies can be computed for a wide range of values of α, m, n. Table 1 contains values of (2) for $\alpha = .05, .01$ and several values of m and n.

For the situation considered here, a power efficiency of $100r\%$ has the quantitative interpretation that the given test based on samples of size m and n has approximately the same power function as the corresponding most powerful test based on samples of size rm and rn. Intuitively the power efficiency of a test measures the percentage of available information per observation which is utilized by that test.

2. Power efficiency derivations. The basic notion of the power efficiency of a significance test is given in [2]. For the present case the problem is to determine the value r such that a most powerful test of the same hypothesis (same significance level) based on rm and rn sample values will have approximately the same power function as the given t-test based on m and n sample values (from $N(a_1, \sigma_1^2)$ and $N(a_2, \sigma_2^2)$ respectively). Here the value of σ_1^2/σ_2^2 is assumed to be known. Then the power efficiency of the given t-test equals $100r\%$.

If the ratio of variances σ_1^2/σ_2^2 is known, the most powerful significance test (one-sided and symmetrical) for the difference of means of the two normal populations is a t-test where the numerator of t is based on the difference of the

TABLE 1

Percentage Power Efficiencies for Certain Values of m and n

$\alpha = .05$

m \ n	4	6	10	15	20	30	50	100	∞
4	79.6	73.5	67.2	63.4	61.4	59.3	57.6	56.2	54.9
6		86.9	82.9	80.2	78.7	77.0	75.5	74.2	72.9
10			92.6	90.9	89.8	88.6	87.3	86.2	85.0
15				95.2	94.4	93.5	92.5	91.5	90.3
20					96.4	95.7	94.9	94.0	92.9
30						97.7	97.1	96.4	95.3
50							98.6	98.1	97.2
100								99.3	98.6
∞									100.0

$\alpha = .01$

m \ n	6	8	10	15	20	30	50	100	∞
6	74.9	70.2	66.7	61.2	57.9	54.3	51.1	48.6	45.9
8		81.3	78.8	74.7	72.1	69.1	66.3	63.9	61.4
10			85.3	81.9	79.8	77.2	74.7	72.5	69.9
15				90.4	88.9	87.0	85.0	83.1	80.7
20					92.9	91.4	89.8	88.1	85.8
30						95.3	94.1	92.8	90.7
50							97.2	96.3	94.5
100								98.6	97.3
∞									100.0

two sample means while the denominator is based on the square root of a function of the sample values and σ_1^2, σ_2^2 which has a χ^2-distribution with $m + n - 2$ degrees of freedom [1, p. 43]. Thus the problem is that of comparing the power functions of two t-tests.

As stated in section 1, it is sufficient to consider one-sided tests. We find, using a modification of the normal approximation to the power function of a one-sided t-test given in [3], that Scheffé's one-sided t-test for the Behrens-Fisher problem and the corresponding most powerful one-sided test (σ_1^2, σ_2^2 known) have approximately the same power function when r is chosen so that

$$K_\alpha - \delta\sqrt{r}\{1 - K_\alpha^2/2[(m + n)r - 2]\}^{1/2} \doteq K_\alpha - \delta[1 - K_\alpha^2/2(m - 1)]^{1/2},$$

where α is the significance level of the tests, K_α is the value of the standardized normalized deviate exceeded with probability α, and δ is a function of m, n, a_1, a_2, σ_1^2, σ_2^2 and the given hypothetical value of $a_1 - a_2$ being tested. This condition for the approximate equality of the power functions is reasonably accurate for the following cases: $\alpha = .05$, $m \geq 4$; $\alpha = .025$, $m \geq 5$; $\alpha = .01$, $m \geq 6$; $\alpha = .005$, $m \geq 7$. The accuracy of the approximation increases as m increases.

Hence a value of r such that the two power functions are approximately equal is determined by the equation

$$r\{1 - K_\alpha^2/2[(m + n)r - 2]\} = 1 - K_\alpha^2/2(m - 1). \tag{1}$$

Let

$$A = A(m, \alpha) = 1 - K_\alpha^2/2(m - 1).$$

Then solving (1) for the appropriate root yields

$$r = \frac{1}{2(m + n)}\{2 + (m + n)A + K_\alpha^2/2 + \sqrt{[2 + (m + n)A + K_\alpha^2/2]^2 - 8(m + n)A}\}.$$

Thus the power efficiency of Scheffé's one-sided t-test solution to the Behrens-Fisher problem, for the case in which the ratio of the variances is also known, is approximately equal to

$$\frac{50}{(m + n)}\{2 + (m + n)A + K_\alpha^2/2 + \sqrt{[2 + (m + n)A + K_\alpha^2/2]^2 - 8(m + n)A}\}\%$$

for suitable values of α and m.

REFERENCES

[1] HENRY SCHEFFÉ, "On solutions of the Behrens-Fisher problem based on the t-distribution," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 35–44.

[2] JOHN E. WALSH, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 64–81.

[3] N. L. JOHNSON AND B. L. WELCH, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), p. 376.

A NOTE ON FISHER'S INEQUALITY FOR BALANCED INCOMPLETE BLOCK DESIGNS

By R. C. Bose

Institute of Statistics, University of North Carolina

1. An experimental design in which v varieties or treatments are arranged in b blocks, is called a *balanced incomplete block design* if

(i) Each block has exactly k treatments ($k < v$) no treatment occurring twice in the same block.

(ii) Each treatment occurs in exactly r blocks.

(iii) Any two treatments occur together in exactly λ blocks.

It is easy to see that the parameters v, b, r, k, λ of the design satisfy the relations

$$bk = vr \tag{1.0}$$

$$\lambda(v - 1) = r(k - 1). \tag{1.1}$$

Also it is readily seen that

$$r > \lambda \tag{1.2}$$

for otherwise with any given treatment every other treatment would occur in every block. This would make $k = v$, and the design would become a 'randomised block design'.

Fisher (1940), showed that a necessary condition for the existence of a balanced incomplete block design with v treatments and b blocks is

$$b \geqq v. \tag{1.3}$$

It is the object of this note to give a very simple proof of Fisher's inequality.

2. Consider a balanced incomplete block design with parameters

$$v, b, r, k, \lambda \tag{2.0}$$

and let

$$n_{ij} = 1 \text{ or } 0 \tag{2.1}$$

according as the ith treatment does or does not occur in the jth block. Clearly

$$\sum_{j=1}^{b} n_{ij}^2 = r \tag{2.2}$$

$$\sum_{j=1}^{b} n_{ij} n_{i'j} = \lambda \qquad (i \neq i'). \tag{2.3}$$

If possible let $b < v$. Consider the $v \times v$ matrix

$$N = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1b} & 0 & \cdots & 0 \\ n_{21} & n_{22} & \cdots & n_{2b} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ n_{v1} & n_{v2} & \cdots & n_{vb} & 0 & \cdots & 0 \end{bmatrix} \tag{2.4}$$

where the last $v - b$ columns of N consist of zeros. It follows from (2.2) and (2.3) that

$$NN' = \begin{bmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots \\ \lambda & \lambda & \cdots & r \end{bmatrix} \tag{2.5}$$

where N' denotes the transpose of N.

$$\det (NN') = |r + \lambda(v - 1)| \ (r - \lambda)^{v-1} \tag{2.6}$$

But $$= kr(r - \lambda)^{v-1} \quad \text{from (1.1).}$$

$$\det (NN') = \det N \det N' = 0. \tag{2.7}$$

This makes $r = \lambda$, and contradicts (1.2). Hence the assumption $b < v$ is wrong, and we must have

$$b \geq v \tag{2.8}$$

REFERENCES

[1] R. A. Fisher, "An examination of the different possible solutions of a problem in incomplete blocks," *Annals of Eugenics*, London, Vol. 10 (1940), pp. 52-75.

[2] F. Yates, "Incomplete randomised blocks," *Annals of Eugenics*, London, Vol. 7 (1936), pp. 121-140.

ABSTRACTS OF PAPERS

(Presented September 1, 1949 at Boulder at the Twelfth Summer Meeting of the Institute)

1. Structure of Statistical Elements.

Duane M. Studley, Foundation Research, Colorado Springs, Colorado.

Research in logical semantics and in practical elementation has set forth the proposition that all words and ideas have set form. As a consequence of this universal proposition all notions and conceptions in statistics should be accessible to set-theoretic analysis and interpretation. This paper explains the results of a preliminary analysis performed on statistical notions and conceptions with a view to a proper organization of definitions and conceptions which will, it is hoped, make possible a better and simpler construction of statistics from a system of basic notions.

2. On the Relative Efficiencies of BAN Estimates.

LEO KATZ, Michigan State College, East Lansing, Michigan.

J Neyman, in the Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1949, proved that χ^2 minimum estimates with either of two alternative definitions of χ^2 are efficient, as also are the maximum likelihood estimates. He also raised the question whether some of these estimates were better than others. This paper bears on that question. In making χ^2 minimum estimates, it is often necessary to avoid small frequencies by grouping together at least one tail of the distribution. It is with respect to the parameters of these *modified* distributions that the χ^2 estimates are efficient Define *relative efficiency* in these circumstances as the ratio of the variance of an efficient estimator in the unmodified case to that of one in the *modified* case. It is shown that, except for a rectangular probability law, the relative efficiency <1 and, further, it decreases as the tail grouping is made wider. Formulae are given for the relative efficiencies of χ^2 minimum estimators for Binomial and Poisson probability laws and some representative values computed to exhibit these effects.

3. Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix.

JACK SHERMAN and WINIFRED J. MORRISON, The Texas Company Research Laboratories, Beacon, New York.

A simple computational procedure is derived for obtaining the elements b'_{ij} of a nth order matrix (B') which is the inverse of (A'), directly from the elements b_{ij} of a matrix (B) which is the inverse of (A), when (A') differs from (A) only in the elements of one column, say the Sth column. The equations which form the basis of the computation are.

$$b'_{Sj} = \frac{b_{Sj}}{\sum_{r=1}^{n} b_{Sr} a'_{rS}}, \qquad j = 1, 2, \cdots n.$$

$$b'_{ij} = b_{ij} - b'_{Sj} \sum_{r=1}^{n} b_{ir} a'_{rS}, \qquad \begin{matrix} i = 1, 2, \cdots S-1, S+1, \cdots n \\ j = 1, 2, \cdots n. \end{matrix}$$

Analogous equations are derived for the case that A and A' differ in the elements of a given row rather than a column.

4. On the Problem of Optimum Classification.

PAUL G. HOEL, University of California at Los Angeles.

Let f_i, $(i = 1, 2, \cdots, k)$, be the probability density function of population i and let p_i be the probability that population i will be sampled. Assume certain differentiability conditions and moment properties. Then, for known parameters, the probability of a correct classification will be maximized by choosing the region M_i, which corresponds to classifying into population i, as that part of variable space where $p_i f_i \geq p_j f_j$, $(j = 1, 2, \cdots, k)$. If the parameters are unknown, an asymptotically optimum set of estimates will be given by the set that minimizes a certain form in the covariances. Among uncorrelated estimates, maximum likelihood estimates are seen to be asymptotically optimum.

If weight functions, W_{ij}, are introduced and the expected value of the loss is minimized, the same methods of proof show that the region M_i becomes that part of variable space where $\sum_{r=1}^{k} p_r f_r (W_{rj} - W_{ri}) \geq 0$, $(j = 1, 2, \cdots, k)$, and that the criterion for an asymptotically optimum set of estimates is of the same form as the preceding criterion.

5. Optimal Linear Prediction of Stochastic Processes whose Covariances are Green's Functions. C. L. Dolph and M. A. Woodbury, University of Michigan, Ann Arbor.

A method of unbiased, minimal variance, linear prediction is developed for problems similar to those of prediction and filtering treated by Wiener. It differs from these in that, the unbiased condition is imposed, only a finite part of the past is employed, and no stationary assumption is used. It is shown that the special stationary case discussed by Cunningham and Hund, "Random Processes in Problems of Air Warfare" (*Supp. Journal Royal Stat. Soc.*, 1946) succeeds because the correlation function, $e^{-k|t-s|}$, well known to that of the process defined by the Langevian equation, is the Green's function of the homogeneous differential equation formed by letting the adjoint differential operator of the Langevian equation operate on the operator of this equation. This relationship is shown to persist for any physically stable linear differential equation driven by "white noise." The well-known equivalence between integral and differential equations is then extended by use of Stieltjes integrals and used to effect the solutions of the integral equations of the first kind which yield the "optimum" linear prediction. The nonstationary example consisting of purely random motion about a mean linear path in the presence of radar type errors is treated in detail.

6. The Integral of the Gaussian Distribution over the Area Bounded by an Ellipse. H. H. Germond, RAND Corporation, Santa Monica, California.

This paper describes the preparation of tables from which to obtain the integral of a bivariate Gaussian distribution over the area of an ellipse. The center of the ellipse need not coincide with the mean of the Gaussian distribution, nor need the axes of the ellipse have any special orientation with respect to the Gaussian distribution.

7. Theorems on Convergency of Compound Distributions with Symmetric Components. (*By title*) Maria Castellani, University of Kansas City.

The purpose of this paper is to present some results obtained when operations of convolution in R_1 are concerned with a specific family of distributions. The compound distribution $K(x) = F(x) * G(x)$ is here obtained combining any d.f. $F(x)$ with a d.f. $G(x)$ under the restriction of symmetry, i.e., $G(x+h) + G(x-h) = 1$ for any $h > 0$.

A generalization of Cantelli's Inequalities will enable us to write a preliminary theorem on the following upper and lower bounds:

$$F(a-h) - 2\int_h^\infty dG(y) < K(a) < F(a+h) + 2\int_h^\infty dG(y),$$

$$K(a-h) - 2\int_h^\infty dG(y) < F(a) < K(a+h) + 2\int_h^\infty dG(y),$$

where a is any point in R_1 and $h > 0$.

The theorem is derived assuming the Stieltjes Integral,

$$K(a) = \int_{-\infty}^{+\infty} F(a-y)\, dG(y),$$

is taken as a sum of three integrals connected with three convenient intervals $(-\infty, -h)$, $(-h, h)$, $(h, +\infty)$. When the symmetric component of the convolution is a member of a fam-

ily of normal distributions such as $G_\alpha(x) = \frac{2}{\sqrt{\pi}} \int_{\alpha}^{x} e^{-\alpha^2 y^2}\, dy$, where α is an arbitrary parameter, the use of Cantelli's Inequalities gives

$$K_\alpha(a - h) - K_\alpha(a) - \frac{2}{\sqrt{\pi}} \int_{\alpha h}^{\infty} e^{-u^2}\, du < F(a) - K_\alpha(a) < K_\alpha(a + h) - K_\alpha(a) + \frac{2}{\sqrt{\pi}} \int_{\alpha h}^{\infty} e^{-u^2}\, du,$$

where $K_\alpha(x) = F(x) * G_\alpha(x)$.

The d.f. $K_\alpha(x)$ is a continuous point function in R_1, with a fr. f. $\gamma(x)$ which is everywhere uniformly continuous. For an arbitrarily small $\eta > 0$, a convenient small h and large α may be found which will enable us to prove the following two theorems:

THEOREM 1. Given any d.f. $F(x)$ in R_1, there exists a convenient continuous d.f. $K_\alpha(x)$ which for $\alpha \to \infty$ converges asymptotically and uniformly almost everywhere to the given d.f. $F(x)$.

THEOREM 2: Given any d.f. $F(x)$ in R_1, there exists in any continuity bordered interval a convenient uniformly convergent series of continuous functions which asymptotically approach the given $F(x)$.

8. Partial Sums of the Negative Binomial in Terms of the Incomplete Beta-Function. (*By title*) JULIUS LIEBLEIN, Statistical Engineering Laboratory, National Bureau of Standards.

In acceptance sampling a certain size sample is taken at random from a lot of items and the lot is accepted if the number of defective items do not exceed a predetermined number characteristic of the sampling plan. The Statistical Engineering Laboratory has been studying the probabilities that a decision to accept or reject can be made before the sample is completely inspected. Such probabilities are found to involve certain sums apparently not previously treated. In this note the author proves a simple identity connecting these sums which greatly facilitates their computation and shows how they may be written in terms of the well known incomplete beta-function of Karl Pearson, for which extensive tables are available.

9. Large Sample Tests and Confidence Intervals for Mortality Rates. (*By title*) JOHN E. WALSH, RAND Corporation, Santa Monica, California.

In computing mortality rates from insurance data, the unit of measurement used is frequently based on number of policies or amount of insurance rather than on lives. Then the death of one person may result in several units of "death" with respect to the investigation; moreover, the number of units per individual may vary noticeably. Thus the usual large sample methods of obtaining significance tests and confidence intervals for the true value of the mortality rate are not applicable in these situations. If the number of units associated with each person in the investigation were known, accurate large sample results could be obtained; however, determination of the number of units associated with each individual would require an extremely large amount of work. This article presents some valid large sample tests and confidence intervals for the mortality rate which do not require much work and are reasonably efficient. The procedure followed consists in first dividing the risks into twenty-six subgroups on the basis of the first letter of the last name of the person insured. Some of the groups are then combined until 10 to 15 subgroups yielding approximately the same number of units are obtained. The fraction consisting of the total number of units paid divided by the total number of units exposed is computed

for each subgroup. Asymptotically the resulting observations represent independent observations from continuous symmetrical populations with common median equal to the true value of the rate of mortality. Tests and confidence intervals for the rate of mortality are obtained by applying the results of the paper "Some Significance Tests for the Median which are Valid Under Very General Conditions" (*Annals of Math. Stat.*, Vol. 20 (1949), pp. 64-81) to these observations.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Fred C. Andrews will be a teaching assistant in the Statistical Laboratory, Department of Mathematics, University of California for the academic year 1949–1950.

Dr. Joseph Berkson has been promoted to the rank of Professor in the University of Minnesota Graduate School and Mayo Foundation. He continues as Chief of the Division of Biometry and Medical Statistics of the Mayo Clinic.

Mr. Colin R. Blyth is now a research assistant at the University of California, Statistical Laboratory, Berkeley.

Mr. Clyde A. Bridger is now Director of the Section of Statistics and State Registrar of Vital Statistics for the Division of Health of Missouri.

Mr. Loren V. Burns, formerly with the MFA Milling Company at Springfield, Missouri, has been made Vice-President and Technical Director of the Spear Mills, Inc., Kansas City 6, Missouri.

Professor Douglas Chapman, who obtained his Ph.D. in statistics at the University of California, Berkeley, has accepted an appointment as Assistant Professor at the University of Washington in the Department of Mathematics and the Laboratory of Statistical Research.

Dr. Andrew Laurence Comrey, who received his doctor's degree from the University of Southern California last June, has accepted an assistant professorship in the Department of Psychology at the University of Illinois.

Dr. Donald A. Darling has been appointed to an instructorship in the Department of Mathematics, University of Michigan.

Dr. Paul M. Densen resigned his position as Chief of the Division of Medical Research Statistics of the Department of Medicine and Surgery of the Veterans Association as of July 1, 1949 to join the staff of the Graduate School of Public Health, University of Pittsburgh, as an Associate Professor of Biostatistics.

Mr. Amron H. Katz has been promoted to the position of Chief Physicist of the Photographic Laboratory, Engineering Division, Air Material Command, Wright Patterson Air Force Base, Dayton, Ohio.

Associate Professor Louis Guttmann, who had been on leave for two years from the Department of Sociology of Cornell University conducting a research program in Israel, was invited to remain in Israel for another year to direct the

activities of the recently founded Israel Institute of Public Opinion Research. He is serving as Chief Consultant.

Mr. Herne Ernest LaFontant who was attending the University of Michigan during the academic year 1948–1949 working on his doctor's degree, has accepted a position as statistician for the B.T.W. Insurance Co. at Birmingham, Alabama.

Assistant Professor Jerome C. R. Li has been promoted to Associate Professor of Mathematics at the Oregon State College, Corvallis, Oregon.

Professor H. B. Mann of Ohio State University has accepted a visiting professorship and research associateship at the Statistical Laboratory at Berkeley, California for the year 1949–1950.

Dr Gottfried E. Noether has been appointed to an instructorship at New York University.

Dr. G. R. Seth has just returned from a trip to England, Sweden, France and India where he visited statistical institutions.

Assistant Professor Andrew Sobczyk has been promoted to Associate Professor of Mathematics at Boston University.

Dr. Zenon Szatrowski, formerly with the Economics Department of Northwestern University, has accepted an associate professorship in the School of Business Administration, University of Buffalo.

Professor Gerhard Tintner has returned to his teaching and research duties at Iowa State College after spending a year at the Department of Applied Economics at the University of Cambridge, England. He gave a course on Econometrics at the University of Cambridge and during his stay in Europe, he lectured on econometric and statistical subjects in Universities at Bristol, Dublin, Hull, Paris, Manchester and Uppsala.

Dr. A. E. R. Westman, Director of the Department of Chemistry, Ontario Research Foundation, left in September, 1949 for England where he is visiting industrial research laboratories and engaging in studies in the Department of Physical Chemistry, Cambridge University. He plans to return in June, 1950.

Word has just been received here of the formation of the New Zealand Statistical Association. The initial meeting was held in August, 1948 at Victoria University College. The officers are: J. T. Campbell, President; I. D. Dick, Secretary. It is planned to hold one formal meeting a year at first with the hope of increasing this later. The main interest in statistical work in New Zealand has been biological, but there is scope for considerable extension to industrial, educational and economic fields and it is hoped the formation of the Association will assist in this extension.

New Members

The following persons have been elected to Membership in the Institute (June 1, 1949 to August 22, 1949)

Al-Doori, Younis A., Student at the University of California, *1915 Henry Street, Berkeley, California.*

Bleber, Robert A., A.B. (Univ. of Calif.) *8418 Richmond Terrace, Richmond, California.*

Bula, Clotilde Angelica, Ph.D., (Univ. of Rosario, Argentina) *Professor*, University of Buenos Aires, *Riopa 2681, Obras Pcia de Buenos Aires, Argentina*

Dalziel, Edwin R., Ph.D. (Univ. Edinburgh) *Assistant Master*, Palmerston North Technical School, Palmerston North, New Zealand

Douglas, James B., Dip. Ed. (Melbourn Univ.) *Lecturer in Mathematics*, Newcastle Technical College, Tighe's Hill 2N, N.S.W., Australia.

Hartley, Herman O., Ph.D. (Cambridge Univ.) *Lecturer in Statistics*, Department of Statistics, University College, London, W.C.1, England

Immel, Eric R., M.A. (Queen's Univ., Kingston, Canada) *Teaching Assistant and Graduate Student*, Department of Mathematics, University of California at Los Angeles, Los Angeles, California.

Kelly, John P., *Senior Technical Engineer*, Carbide and Carbon Chemical Corporation, Oak Ridge, Tennessee, *P.O. Box 478, Norris, Tennessee*

Parel, Cristina P., M.S. (Univ. of Michigan) *Instructor*, Department of Mathematics, University of the Philippines, Manila, P.I.

Philipson, Carl O., D.Sc. (Univ. of Stockholm) *Actuary of Folket Samarbete*, *Ynggvagen 5, Djursholm, Sweden.*

Porter, Robert A., Ph.D. (N.C. State College, Raleigh, N.C.) *Senior Mathematician*, University of Chicago, *17113 Longfellow Avenue, Homewood, Illinois*

Rippe, Dayle D., M.A. (Univ. of Nebr.) *Student, Teaching Fellow*, Department of Mathematics, University of Michigan, *1048 Woburn Court, Willow Run, Michigan.*

Rogers, Robert L., A.B. (Univ. of Calif.) *Student at University of California*, *Route 2, Box 74, Denio Avenue, Gilroy, California.*

Roy, Samarendra N., M.Sc. (Calcutta Univ.) *Head of Department of Statistics*, Calcutta University and *Assistant Director*, Indian Statistical Institute (now on leave) *P.O. Box 168, Chapel Hill, North Carolina.*

Savey, Rosemary, M.B.A. (Univ. of Wisc.) *Graduate Assistant and Student*, University of Wisconsin, *2513 Norwood Place, Madison 5, Wisconsin.*

REPORT ON THE BOULDER MEETING OF THE INSTITUTE

The Twelfth Summer Meeting of the Institute of Mathematical Statistics was held at the University of Colorado, Boulder, Colorado, Monday, August 29 through Thursday, September 1, 1949. The meeting was held in conjunction with the summer meetings of the American Mathematical Society, the Mathematical Association of America, and the Econometric Society. The meeting was attended by the following 79 members of the Institute:

S. P. Agarwal, R. L. Anderson, T. W. Anderson, V. L. Anderson, K. J. Arnold, E. W. Barankin, C. A. Bennett, Agnes Berger, E. E. Blanche, A. H. Bowker, J. C. Brixey, Jean Bronfenbrenner, J. H. Bushey, H. C. Carver, Herman Chernoff, K. L. Chung, A. G. Clark, E. P. Coleman, E. L. Crow, J. H. Curtiss, W. J. Dixon, J. L. Doob, Aryeh Dvoretzky, H. P. Evans, W. D. Evans, W. T. Federer, William Feller, C. H. Fischer, J. S. Frame, T. C. Fry, H. M. Gehman, H. H. Germond, R. E. Greenwood, H. T. Guard, P. R. Halmos, J. L. Hodges, P. G. Hoel, Harold Hotelling, J. M. Howell, C. C. Hurd, C. A. Hutchinson, Irving Kaplansky, Leo Katz, H. S. Konijn, T. C. Koopmans, G. M. Kuznets, H. D. Larsen, D. H. Leavens, S. B. Littauer, H. B. Mann, Jacob Marschak, F. J. Massye, Dorothy J. Morrow, Jerzy Neyman, M. L. Norden, J. I. Northam, E. G. Olds, R. P. Peterson, G. B. Price, Mina S. Rees, P. R. Rider, F. D. Rigby, Herman Rubin, L. J. Savage, Elizabeth R. Scott,

I. E. Segal, Esther Seiden, Jack Sherman, W. B. Simpson, Milton Sobel, D. M. Studley, B. R. Suydam, A. G Swanson, James Templeton, R. M. Thrall, J. W. Tukey, Abraham Wald, John Wishart, S. S. Wilks.

The Monday afternoon session was devoted to invited addresses with Professor Leonard J. Savage of the University of Chicago presiding. The attendance was approximately fifty. Professor J. L. Hodges of the University of California presented a paper, *Some Problems in Point Estimation*, and Professor W. T. Federer of Cornell University presented *A Comparison of the Proportionality of Covariance Matrices*.

On Tuesday Morning the Institute, the Mathematical Association of America, and the Econometric Society held a joint symposium on *Mathematical Training for Social Scientists*. Professor Jacob Marschak of the Cowles Commission for Research in Economics presided. The attendance was approximately one hundred fifty. The participating speakers were: Professor R. L. Anderson of North Carolina State College; Professor T. W. Anderson of Columbia University; Professor G. C. Evans of the University of California, Professor F. L. Griffin of Reed College; Professor Harold Gulliksen of Educational Testing Service, Professor William Jaffé of Northwestern University, Professor Harold Hotelling of the University of North Carolina; and Professor G. M. Kuznets of the University of California. At the end of the session the following resolution was adopted by those in attendance at the meeting:

> Members of the Mathematical Association of America, the Institute of Mathematical Statistics, and the Econometric Society assembled in a joint session in Boulder, Colorado, on August 30, 1949, are of the opinion that officers of these societies should study the need for better mathematical training of social scientists, and the ways and means to improve mathematical preparation of social scientists, and that such a study may be most effectively conducted by a joint committee, possibly in co-operation with other interested societies, and in close touch with the Social Science Research Council, the National Research Council, or other national bodies concerned with general education and research. It is suggested that this committee report the results of its deliberations at the next joint meeting of the original participating societies

The two joint sessions of the Institute and the Econometric Society were devoted to a Symposium on *Statistical Inference in Decision Making*. Professor Jerzy Neyman of the University of California presided on Tuesday afternoon. The attendance was approximately eighty. Professor Aryeh Dvoretzky of Hebrew University, Jerusalem presented *Decision Problems* and Professor Abraham Wald of Columbia University presented *Some Recent Results in the Theory of Statistical Decision Functions*. On Wednesday Morning, under the chairmanship of Professor Wald and an attendance of approximately seventy-five, the following papers were presented: *Remarks on a Rational Selection of a Decision Function* by Professor Herman Chernoff of the Cowles Commission for Research in Economics; *Psychological Probabilities* by Professor Leonard J. Savage; and *Complete Classes of Decision Functions for Some Standard Sequential and Non-sequential Problems* by Milton Sobel of Columbia University.

On Thursday Morning the Institute and the American Mathematical Society held a joint session for contributed papers with Professor P. R. Rider of Washington University presiding. The attendance was approximately seventy-five. The following papers were presented:

1. *Structure of Statistical Elements*
   Mr. Duane M. Hurley, Foundation Research, Colorado Springs.
2. *On the Relative Efficiencies of BAN Estimates*
   Professor Leo Katz, Michigan State College
3. *Adjustments of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix*
   Dr. Jack Sherman and Miss Winifred J. Morrison, The Texas Company Research Laboratories, Beacon, New York.
4. *On the Problem of Optimum Classification*
   Professor Paul G. Hoel, University of California at Los Angeles
5. *Optimal Linear Prediction of Stochastic Processes whose Covariances are Green's Functions*
   Professor C. L. Dolph and Dr. M. A. Woodbury, University of Michigan.
6. *The Integral of the Gaussian Distribution over the Area Bounded by an Ellipse.*
   Dr. H. H. Germond, Rand Corporation, Santa Monica, California
7. *Theorems on Convergency of Compound Distributions with Symmetric Components.* (By title)
   Dr. Maria Castellani, University of Kansas City
8. *Large Sample Tests and Confidence Intervals for Mortality Rates.* (By title)
   Dr. J. E. Walsh, Rand Corporation, Santa Monica, California
9. *Partial Sums of the Negative Binomial in Terms of the Incomplete Beta function.* (By title)
   Dr. Julius Lieblein, National Bureau of Standards.

On Thursday afternoon Professor Jerzy Neyman presented the Second Rietz Memorial Lecture on *Consistent Estimates of the Linear Structural Relation in the General Case of Identifiability*. Professor Harold Hotelling presided and the attendance was approximately fifty. Dr. R. P. Boas, Jr. of Mathematical Reviews presented an invited address *The Representation of Probability Distributions by Charlier Series*.

The Institute sponsored a beer party on Tuesday evening and on Thursday evening a fry was held on Flagstaff Mountain.

HARRIS T. GUARD
Assistant Secretary